



武汉大学

本科生课程讲义及知识总结

课程名称：网络 GIS

开课学院：遥感信息工程学院

开课时间：2020-2021 年度第一学期

2020. 武汉大学

目录

第一章 · 绪论.....	3
➤ 相关概念.....	3
第二章 · 网络 GIS 基础.....	5
➤ 概述.....	5
➤ 网络 GIS 体系结构.....	6
➤ 网络 GIS 数据组织与管理.....	10
➤ 空间数据共享与安全.....	14
➤ 网络 GIS 通信协议及规范.....	18
➤ 网络 GIS 实现技术.....	20
第三章 · 空间数据网络存储.....	23
➤ 数据存储技术.....	23
➤ 网络存储分类.....	29
➤ 网络存储模式.....	32
➤ 云存储.....	33
➤ 网络 GIS 数据存储实例.....	35
第四章 · 移动 GIS 技术.....	35
➤ 移动 GIS 的组成.....	35
➤ 移动空间数据管理.....	37
➤ 移动 GIS 设计与开发以及 Android 的介绍.....	40
➤ 移动 GIS 的应用.....	42
第五章 · GIS 云.....	43
➤ 云计算.....	43
➤ GIS 云基础.....	44
➤ GIS 数据云.....	46
➤ GIS 分析云.....	47
➤ GIS 云应用.....	48

第一章 · 绪论

➤ 相关概念

1. **GIS (地理信息系统) 的概念**: 是一种采集、传输、存储、管理、处理、表达和使用地理空间数据的计算机系统。
2. **GIS 的 5 个基本功能**: 数据输入、数据编辑、数据存储与管理、空间查询与空间分析、图形输出与交互操作。**【空间查询**: 根据特定条件对空间数据进行查询和检索, 包括位置查询、属性查询和拓扑查询**】**
3. **GIS 的主要特点 (从 5 个方面)**:
 - 1) **空间数据组成**: 包括空间数据和属性数据; 数据量大且类型复杂; 来源广泛且非结构化。
 - 2) **特有的空间分析能力**: 利用空间解析模型和应用分析模型分析空间数据实现快速的空间定位检索和复杂的查询功能。
 - 3) **强大的图形处理与表达能力**。
 - 4) **辅助决策支持**: 空间分析提供空间模拟和空间决策支持。
 - 5) **智慧城市/数字城市地理信息公共服务平台**: 充当地理信息公共服务平台的角色。
4. **World Wide Web**, 通常称为万维网, 缩写为 WWW。基本概念有超文本、超媒体、超文本标记语言 (HTML)、超文本传输协议 (HTTP)。
5. **分布式计算技术的发展**:
 - 1) **集中式**: 基于主机的计算模式、基于服务器的计算模式;
 - 2) **分布式**: 客户端/服务器 (C/S)、浏览器/服务器 (B/S)、多层分布式计算模式、Web Service 计算模式、网格计算模式、网络计算模式、云计算、边缘计算。
6. **数字地球**: 用数字化的方法, 将地球、地球上的活动以及地球环境进行虚拟化后装入电脑中, 通过高速网络实现空间信息交换和共享, 使之最大限度地为人类的生存、可持续发展和日常的工作学习服务提供操作环境和工具。
7. **智慧地球**: 智慧地球是数字地球的智慧化。
8. **数字城市的局限**:
 - 1) 信息获取手段的自动化程度低, 缺乏实时更新, 更新准确度低;
 - 2) 难以实现城市直接相关的感知设备的接入;

- 3) 标准、技术、观念和部门利益等导致信息资源整合之后,融合共享困难,烟囱效应突出;
- 4) 应用效率和智能化程度低,网上“有数据找不着,数据多用不了”
9. **智慧城市**:是在**数字城市**基础上,通过各种传感器构成的**物联网**对现实城市进行**透彻感知和全面互联互通**,获取城市各种信息。
10. **智慧水利**:以**水利信息化**为核心,以**数字化、网络化、智能化**为特征,是**数字水利**的延伸和深化,是综合治理水资源、水生态、水环境、水灾害四大水问题的重要工程。
11. **供应链**:是一个由供应商、制造商、分销商和终端用户所构成的物流网络,也是一个功能网链。
12. **供应链管理**:以**用户**为中心,以用户需求为导向,以优化网链库存成本为目标,覆盖率原材料和零部件从采购、供应到产品制造、运输、仓储、销售各个环节,包含了**物流、商流、信息流和资金流**全过程管理,是一种集成化的管理模式。
13. **应用 GIS 制作的军用电子地图**,较以往的军用地图具有**精度高、信息量大、可编辑性强、操作简便、便于携带**等特点。
14. **位置服务 (LBS, location based services)**:是 GIS 新兴发展方向,整合**Internet、无线通信、移动定位与 GIS**等技术于一体,提供与位置相关的增值服务,其核心技术是**无线通信技术、移动互联网技术。导航与定位技术和 GIS 技术**。
15. **众源地理数据**:是一种以互联网为传输平台的开放的地理空间数据,主要包括**公共版权数据、GNSS 数据、网络用户自发上传的地理空间数据**等多源地理空间数据。
16. **网络化 GIS (网络 GIS)**:以网络为平台的 GIS,在网络环境下为各种地理信息相关应用提供 GIS 的基本功能、分布式计算和空间数据管理的空间信息管理系统。
17. **GIS 网络化的 5 方面优点**:①拓宽了 GIS 的应用领域和服务范围;②平台独立性,为用户提供透明的操作方式;③支持多源异构数据,提高了数据利用价值;④大规模降低系统的构建成本;⑤时效性增强。
18. **大数据的“5V”特征**:①**内容丰富、形式多样 (variety)**;②**容量巨大 (volume)**;③**数据的处理速度快 (velocity)**;④**数据的正确性 (veracity)**;⑤**数据价值高 (value)**。
19. **高性能并行计算的计算平台有两种实现方式**:①**紧耦合的大型机和巨型机**;②**松耦合的分布式计算机系统**。

第二章 · 网络 GIS 基础

➤ 概述

1. **基于 Web 的 GIS 的概念：**是 GIS 技术和 WWW（World Wide Web，万维网）技术的有机结合，是 Internet 或 Intranet 环境下的一种传输、存储、处理、分析、显示与应用地理空间信息的计算机系统。（WebGIS 是网络 GIS 的主要代表）
2. **传统 GIS 的不足：**①数据互操作性较差；②GIS 数据共享能力弱；③数据冗余严重；④空间分析能力有限；⑤成本高昂；⑥应用拓展困难。
3. **网络 GIS 的特点：**
 - 1) **数据特点：**①采集手段多元；②数据地理分布（空间数据本身的地理分布性+空间数据存储地域的分布性）；③多元、异构、海量（来源广泛、结构多样）；④表现形式多样化（专用浏览器、通用 Web 浏览器、通用浏览器加特定插件）。
 - 2) **数据特点：**①用户分布广泛；②用户类别多样、需求各异；③用户访问随机性大（环境变化容易导致多用户并发访问系统等）。
 - 3) **系统特点：**①基于 Internet/Intranet 标准（以标准的浏览器为客户端，通过 TCP/IP、HTTP、WAP 等协议访问任何地方的通检数据）②分布式体系结构（客户端服务器相互分离、协同工作）；③多与平台无关（网络 GIS 的分布性和多用户特点决定了它必须具备较强的跨平台性能，适用于异构系统，对客户端的软硬件没有特殊要求）；④成本低廉、操作简单；⑤支持地理分布存储的多元数据。
4. **网络 GIS 的功能：**①网络存取；②网络计算（通过选择不同的计算模式调动地理分布的各个节点的资源，实现协同计算，提高空间数据处理效率）；③查询检索与联机处理；④空间数据可视化；⑤空间模型分析与服务（是区分 WebGIS 的最主要特征）；⑥Web 资源共享；⑦空间数据分布。
5. **网络 GIS 对于推动 GIS 技术发展和进步的主要作用：**
 - 1) 促使 GIS 走向网络化、大众化，使 GIS 走近人们的生活、工作和学习中；
 - 2) 加快了空间数据获取和分发速度；
 - 3) 使空间分析和空间信息服务无处不在、无时不有；
 - 4) 为各种智能应用提供地理信息公共服务平台和基础地理空间框架；
 - 5) 使 GIS 与其他软件的集成变得容易，推动 GIS 走向纵深和宽广。

6. 应用服务模式:

- 1) 空间数据下载: 将 GIS 的原始空间数据以 Web FTP 方式从服务器端下载到客户端, 供客户端存储、备份或处理。
- 2) 静态地图显示: 在线浏览服务模式, 为客户端提供能够在线浏览的静态地图图像
- 3) 动态地图浏览: 用户能在浏览地图时还能与之交互;
- 4) 空间元数据查询: 通过发布空间元数据的方式进行空间数据的网络发布, 使用户能方便、及时地了解所关注的空间数据情况, 并通过适当途径得到满足要求的空间数据;
- 5) 数据预处理: 在数据传输之前对原始数据进行一定程度的预处理。

➤ 网络 GIS 体系结构

1. GIS 体系结构大致经历了**单机结构 GIS**和**网络 GIS**两个发展阶段。
2. 与传统 GIS 相比, 数据、系统、用户三要素在**网络 GIS**中构成了**4 种分布: 采集分布、存储分布、处理分布、应用分布**。
3. 在 Internet 支持下, 根据 **TCP/IP** 和 **HTTP** 协议, 网络 GIS 把**支持标准的 HTML(超文本标识语言)**的浏览器作为统一客户端(即 WebGIS 浏览器), 其核心是将 GIS 的功能嵌入到满足 **HTTP** 和 **TCP/IP** 标准的 Internet 应用体系中, 实现 Internet 环境下地理空间信息的有效管理、处理、分析和服

4. 按照逻辑结构, 复杂应用层序可分为 4 层:

【4 个逻辑清晰地表明了各层之间的**数据流动关系**】

- 1) **表示逻辑**: 主要负责前端用户界面;
- 2) **业务逻辑**: 主要为用户提供所需的业务服务;
- 3) **事务逻辑**: 主要为用户业务提供公共或特定的服务, 并负责应用程序访问数据的安全性和完整性;
- 4) **数据逻辑**: 主要负责数据库的存取和管理



图2-1 四层逻辑结构 (新图)

5. 两层体系结构（客户端——服务器）：

- 1) 两层体系结构把网络 GIS 分成客户机和服务器两部分，客户端（客户浏览器）之间通过 Internet 或转网连接，形成客户端/服务器计算模式（C/S 模式）。
- 2) 客户机和服务器相互独立、相互依存。客户机通常是承载最终用户使用的应用软件系统的单台或多台设备；服务器往往是一些高性能计算机，其功能由一组协作的过程或数据库及其管理系统所构成，为客户机提供服务。
- 3) 一般来说，数据逻辑部署在服务器端，表示逻辑部署在客户端，业务逻辑和事务逻辑可以部署在客户端或服务器端。

基于客户机的网络 GIS 体系结构（又称“胖”客户机或“瘦”服务器的网络 GIS）：

- 1) 表示逻辑、业务逻辑和事务逻辑都在客户机一端实现，导致客户端“臃肿”；
- 2) 客户机需要下载动态执行或安装相应客户机 GIS 应用程序，需要请求 GIS 数据及处理工具。

基于服务器端的网络 GIS 体系结构（称“瘦”客户机或“胖”服务器的网络 GIS）：

- 1) 服务器端承担过多的数据处理任务，业务逻辑、事务逻辑、数据逻辑都在服务器一端实现，在响应多个客户机的请求时可能性能跟不上导致数据处理效率低下；
- 2) 客户机仅充当前端的对用户友好的接口；
- 3) 用户在客户机浏览器通过向服务器发送初始化和数据处理、服务请求，服务器收到请求后分析其要求并对请求加以处理，将结果通过网络返回客户机并按适当方式予以显示。

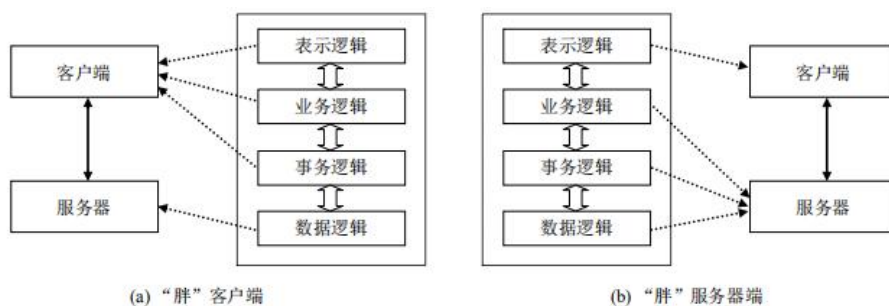


图2-3 两层结构的逻辑实现（新图）

6. 三层体系结构（客户端——中间层服务器——数据服务器）：

- 1) 是基于服务器端体系结构的拓展；
- 2) 三层体系结构将业务逻辑、表示逻辑和数据逻辑适当分开，部署在不同的节点中，从而减轻客户端和服务器的压力，较好地平衡负载。

- 3) 将表示逻辑与 GIS 的处理逻辑分开，使 GIS 的处理逻辑为所有用户共享，从根本上克服两层结构的缺陷
- 4) 应用服务器通常是工作站、小型机或高性能服务器，也可以是服务器集群；中间层通常是工作站或小型机；客户机可以是 PC 级、智能手机或平板电脑。

三层结构的多种形式：

- 1) 中间层（应用服务器）承载全部业务逻辑和事务逻辑；
- 2) 中间层（应用服务器）承载部分业务逻辑和事务逻辑

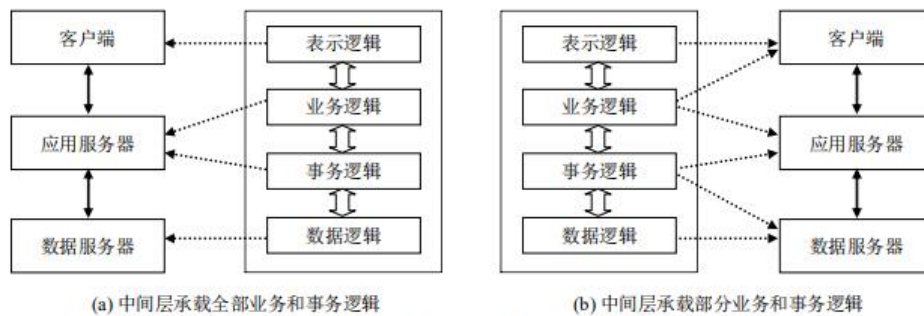


图2-4 三层结构的多种形式（新图）

7. 多层结构：

- 1) 多层结构与三层结构相比，主要是在业务逻辑层增加了更多的逻辑处理单元，以根据不同客户的请求情况分别予以高效处理。
- 2) 典型的多层体系结构网络 GIS：ArcGIS、Autodesk GIS、MapInfo、MapGIS、SuperMap GIS。

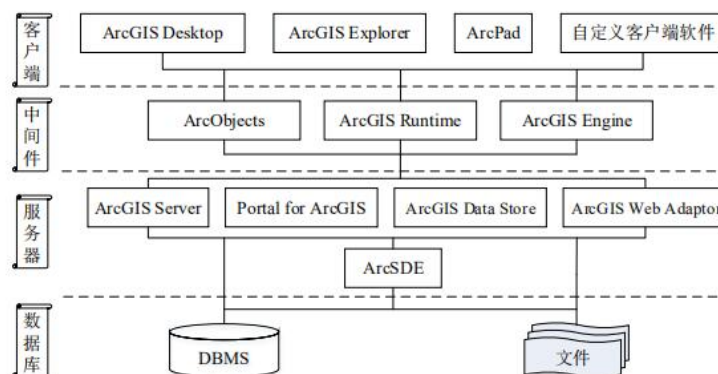


图2-5 ArcGIS系列软件多层体系结构(在原图3-3基础上修改)

8. 基于三层客户/服务器模式的软件平台在逻辑上可以简单地分为用户浏览器、GIS 功能中间件和 GIS 数据存储服务器三部分。
- 1) **数据存储服务器**：负责空间数据和属性数据的存储、管理与维护，以及与 Web 服务器进行数据交互。

- 2) **Web 服务器和 GIS 功能中间件**：是三层结构体系中与客户进行交互的服务器，它负责接收客户端的 Web 请求，根据客户请求，提供相应的 GIS 功能服务，通过与数据存储服务器的数据存取模块的通信实现空间数据库的编辑，并将结果数据返回给客户端。
- 3) **GIS 客户/浏览器**：浏览器形式/插件形式



图2-6 WebGIS三层模型逻辑图（原图3-15）

9. **多服务器系统结构**：多服务器是指物理上相互独立，而逻辑上单一的一组计算机的集群，以统一的系统模式加以调度和管理，为客户端提供高可靠性、快速服务。

工作原理：

- 1) 当用户第一次请求对象服务时，首先向负载调控器发送“获得服务对象实例引用”的请求。
- 2) 负载调控器根据各后台服务器的当前负载状态，从中选择一个合适的服务器，并由该服务器的对象实例管理器分配服务对象实例，同时将该对象的引用返回给客户端。
- 3) 客户端通过获得的对象实例引用指针，完成后续请求操作(不再通过负载调控器转发)。
- 4) 若后台服务器负载过重，则用户建立服务对象实例的请求将被负载调控器暂存到等待队列中进行排队，待服务器空闲时再行处理。

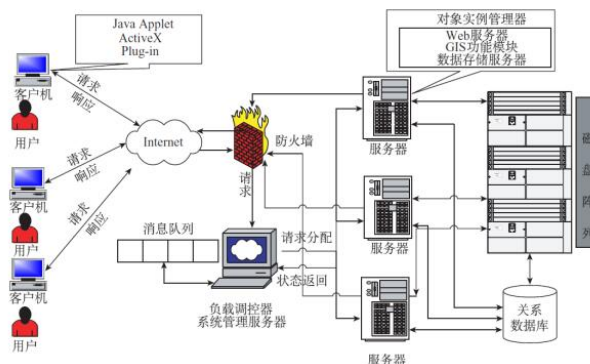
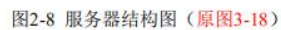


图2-7 多服务器系统结构图（原图3-17）

- 1) **负载调控器**：又称系统管理服务器，主要为多服务器系统提供负载与系统信息的**监控、负载初始化分配、动态资源调度与任务迁移**等功能。
- 2) **对象实例管理器**：对应于系统中的每一台具体的应用服务器，它负责管理该服务器，获取当前服务器运行参数和状态，将这些参数返回给负载调控器。
- 3) **Web 应用服务器**：与对象实例管理器运行在同一台节点服务器上，主要实现业务逻辑功能。



动态负载平衡的主要解决方法有:

- 【调控负载是实现动态平衡的必要措施，主要包括分配新任务请求、负载平衡处理两个阶段】

➤ 网络 GIS 数据组织与管理

- 分层理论是 GIS 数据组织的基本方法。

地理空间数据由若干个图层及相关属性数据组织而成，而每个空间数据图层又以若干个空间坐标的形式存储。（分层组织为：坐标对-空间对象-图层-地图）

1) **地图集**: 地图集是地理数据组织中的顶层信息, 实现对各个地图的管理。主要包

含：地图引用、地图坐标及地图描述等。

2) **图层集**：由多个空间图层构成的能满足一定应用需求的图层集合。包含：组成图层集的图层引用、图层空间索引、图层显示、图层坐标范围等信息。

3) **图层**：由多个具有某些相同或相似特性的同种类型的空间对象构成的集合。包括空间对象标识、描述及几何表示。

3. 矢量数据组织——基于特征的数据组织：

特征的概念：是对地理现象的高度抽象和全面表达，它包括地理现象在**空间、时间和专题**等方面的所有信息，既是一个地理实体又是一个表达对象，既是地理现象也是地理现象的数字化表达。

4. 栅格数据组织的组织方式：

1) **基于像元**：每个像元对应一条记录，包括像元坐标及其各类属性值的编码。

【简单、便于数据扩充和修改，属性查询和面域边界提取速度较慢】

2) **基于层**：层为存储基础，层中以像元为序记录其坐标和对应该层的属性值编码。

【便于属性查询，重复存储每个像元坐标占据不少存储空间】

3) **基于面域**：以层作为存储基础，层中又以面域为单元进行记录，包括面域编号、面域对应该层的属性值编码、面域中所有像元的坐标等。

【有利于面域边界提取，但不同层中像元坐标仍需要多次存储】

5. 栅格数据组织——遥感影像组织：

基本思想：关系数据库与文件组织管理相结合。

1) **分级存储**：按某种规则对影像重采样，使像素变少。原始影像与重采样影像，自下而上构成金字塔结构，位于塔顶影像像素最小，覆盖范围最大，适宜于粗略查看全幅影像；位于塔底影像细节清晰，便于放大处理和查看。

2) **分块存储**：将覆盖范围较大的高分影像分割成覆盖范围适中的多块影像来存储

3) **影像金字塔**：以原始影像为基础重采样。水平尺度分块、垂直尺度分层，生成分辨率由大到小的各层影像数据。

6. 对象-关系型空间数据管理技术：

把**复杂的数据类型作为对象放入关系数据库**中，提供索引机制和操作方法。

对象-关系型空间数据管理系统分为三个层次：**空间应用层、空间数据库管理系统层及商业数据库管理系统层**。

7. 分布式空间数据库管理技术:

分布式数据库: 物理上分布的数据库集合, 由数据库管理软件进行统一管理。数据库集合可看作是松耦合连接的一组节点, 通常不共享任何物理部件, 各节点数据库系统相互独立。

分布式数据库的分类:

- 1) **同构:** 所有节点用共同的数据库管理系统软件, 协同处理用户任务请求;
- 2) **异构:** 不同节点有不同模式和不同数据库管理系统软件, 节点之间互不了解, 彼此间仅提供有限功能合作。

分布式空间数据库管理技术的关键点:

- 1) **采用复制和分片存储一个关系 R。**
- 2) **分布式事务:** 分布式数据库管理系统中对各数据项的访问常常通过事务来完成。
- 3) **提交协议:** 分布式数据库管理系统中必须保证执行的事务是**原子性**的, 即事务必须保证在所有节点上执行产生的**最终结果的一致性**。

8. 并行空间数据库管理技术:

并行数据库系统资源包括: 处理器、主存模块和二级存储。

并行数据库系统按照资源互相作用方式划分体系结构, 三类主要的体系结构分别为共享内存(Shared Memory, SM)、共享硬盘(Shared Hard Disk, SHD)和无共享(Shared Nothing, SN)。

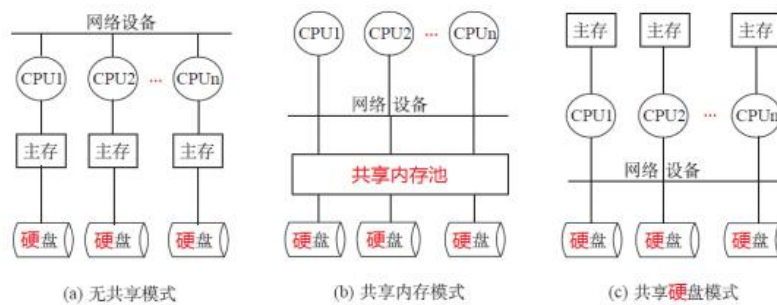


图2-16 并行体系结构 (原图3-10)

- 1) 无共享模式中, 每个处理器只与供其访问的主存和硬盘单元相关。
- 2) 共享内存模式中, 多个处理器通过网络设备互联, 访问一个公共的、系统范围内的主存(内存池)
- 3) 共享硬盘模式中, 各处理器都拥有只能被其本身访问的专用主存, 且所有处理器都可以访问系统中所有的硬盘资源。

9. 自适应数据组织:

基础: 影像数据分块

分块策略: 兼顾存储、传输和客户端内存 3 个方面。

自适应影像分块存储机制包括**数据库自适应切片机制**和**影像的 LOD 模型建立**两部分

1) **自适应切片机制**是指在分片时考虑**数据库存储块大小、带宽和客户端硬件设施**这三个要素的影响,亦即,根据网络中这三个要素的情况综合确定分片策略。

a) **顾及数据库存储块切片:** ①数据读写时,通常以最小存储单元数据块为单位;②数据块大小决定 I/O 索引访问路径,影响读写效率;③数据块较小时,减小块间读写竞争,提高文件访问效率;④数据块较大时,存储单元存储更多数据,额外开销小,但对于多并发读取,会加重索引页块间的竞争。

基于不同大小数据库存储块的影像切片存储耗时(单位: s)

影像切片尺寸 数据库存储块大小	128×128	256×256	512×512	1024×1024	2048×2048
2KB	104.08	45.08	36.62	38.14	40.36
4KB	104.48	40.28	30.17	30.87	32.99
8KB	92.86	33.62	24.04	22.30	23.87
16KB	88.50	33.80	22.76	21.38	21.01

b) **顾及网络带宽的切片:** ①网络带宽足够大时,切片尺寸越大,传输效率越高;②网络状态不稳定,网络带宽难以达到最大值;③设置合适切片大小,可有效减小单次网络传输数据量,减轻传输压力。

基于不同网络带宽的影像切片传输耗时(单位: s)

影像切片尺寸 网络带宽	128×128	256×256	512×512	1024×1024	2048×2048
2Mbps	122.725	52.272	64.83	72.272	149.335
10Mbps	45.568	38.927	24.322	21.188	25.586
64Mbps	29.043	24.988	20.218	18.147	19.863
128Mbps	18.487	14.621	12.208	10.825	10.278

c) **顾及客户端内存容量的切片:** ①内存容量限制,难以将全部空间数据读到内存中处理;②客户端每次操作只涉及原始影像的小部分,无需全部传输;③结合内存缓存技术,通过设置最适宜切片大小,实现海量影像的实时显示,提高影像渲染效率。

基于不同容量客户端内存的影像切片渲染耗时(单位: s)

影像切片尺寸 (每切片数据量) 客户端内存	128×128 (48KB)	256×256 (192KB)	512×512 (768KB)	1024×1024 (3MB)	2048×2048 (12MB)
2GB及平均渲染时间	20.56	8.529	10.353	14.226	22.334
4GB及平均渲染时间	18.046	5.859	3.593	2.489	4.438
8GB及平均渲染时间	13.879	4.272	2.315	1.709	0.754

2) **LOD 模型**是在不影响画面视觉效果条件下,通过逐层简化影像的清晰度来减少影像数据量,进而提高影像的传输和绘制效率。

☆**自适应切片模型**：根据以上 3 个因素得到切片大小。通过为每一个影响因子赋予不同的权重，**自适应切片大小将是这 3 种影响因子下对应的切片大小的加权平均值。**

$$BlockSize = \alpha \times P(T) + \beta \times W(S) + \gamma \times C(N)$$

$P(T)$ ：依客户端内存确定的切片大小； $W(S)$ ：依带宽确定的切片大小； $C(N)$ ：依数据库存储块大小得出的切片大小； α 、 β 、 γ ：对应切片大小的权重，未知数，待求。

例如： $BlockSize = 0.191P(T) + 0.511W(S) + 0.298C(N)$

☆**影像 LOD 模型**：影像数据量庞大，内存有限，难以一次读取到内存，客户端大多数情况下并不需要全部原始影像，仅需其一部分内容，使用影像金字塔的数据存取：

- 1) 确定客户端最佳显示分辨率，在金字塔中找出与最佳分辨率最接近的影像金字塔层。
- 2) 利用建立的空间索引，确定需要调用的影像块。
- 3) 客户端将从服务器端传来的影像块予以拼接显示。
- 4) 构建自适应影像金字塔时，层数一般不宜超过 10 层。

➤ 空间数据共享与安全

1. **基础地理空间数据**：提供基础地图底图服务和空间基准服务的数据。

专题地理空间数据：为满足特定行业需求而生产的与地理空间位置和范围密切相关的业务数据。（通常以基础地理空间数据为基础）

2. **常用 GIS 数据共享方法（异构数据共享方法）**：**数据格式互换模式、数据直接访问模式和数据互操作模式。**

- 1) **数据格式互换模式**：①转换后不能完全准确表达源数据的信息；②数据转换频繁、费时，易产生数据不一致性问题。

例如：ArcGIS 与 Google Earth：通过 ArcGIS 获得 Shape 格式数据，通过 ArcToolbox 将 Shape 转换为 GE 的 KML 文件，KML 文件加载到 GE 中，GE 处理转换后的数据。

- 2) **数据直接访问模式**：①避免了繁琐的数据转换过程，效率高；②无需拥有目标数据格式的宿主软件；③经济实用，适合多源数据共享。

例如：GeoMedia 软件可以直接访问多数软件的空间数据，包括 MGE、Arc/Info、

Oracle Spatial、SQL Server、Access MDB 等数据的直接访问。

- 3) **空间数据互操作**：指通过规范接口自由处理所有种类空间数据的能力和在 GIS 软件平台上通过网络处理空间数据的能力。

空间数据互操作的特点：①遵循互操作规范 OpenGIS；②一个系统可同时支持不同的空间数据格式；③互操作不仅是对**数据**的集成，也是对**处理过程和处理功能**的集成，实现在更高层次上**不同系统、不同环境**之间的互相合作。

3. 分布式空间数据共享：

- 1) **分布式空间数据库**：分布式空间数据库是指空间数据库在**物理上**分布于计算机网络的各个节点中，每个节点拥有一个集中的空间数据库系统，而且都具有自治处理能力。**特点**：①空间数据物理上分布于各节点中，逻辑上统一；②每节点数据具有独立性，拥有集中的空间数据库系统；③节点具有自治处理能力，负责完成本节点局部应用。
- 2) **地理数据分布式计算**：采用分布式计算模式的 GIS 数据和处理功能(应用程序)不仅可以位于一个集中的服务器上，也可以分散在多个地理分布的服务器上。
- 3) **分布式空间数据库的空间数据共享**：分布式空间数据库的空间数据共享是指用户怎样通过 Internet 有目的地从数据服务器透明地获取各种格式的数据。

联邦数据库(Federated DataBaseSystem, FDBS)管理的数据库一体化构想：

目的就是发现**成分模式和联邦模式之间的映射**，并在局部操作和全局操作之间定义这种映射，利用统一数据模型的联邦模式**屏蔽局部模式之间的差异性**。

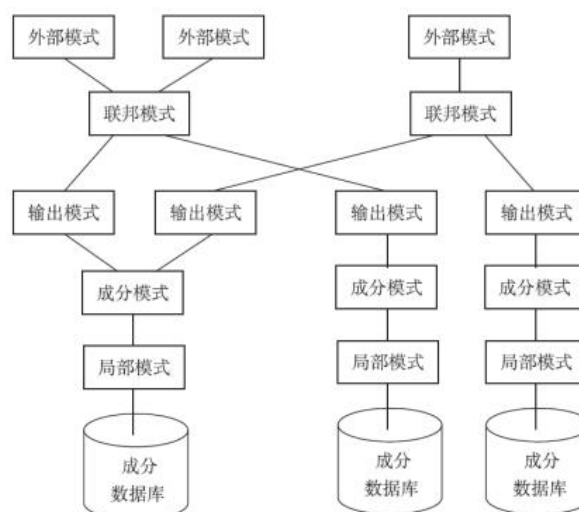
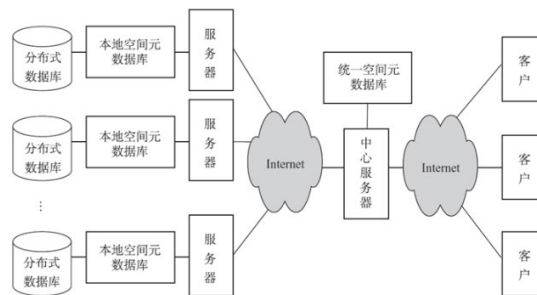


图2-18 联邦数据库的五层模式 (原图3-11)

联邦数据库相关的名词解释：

- ①成分数据库：分布于网络的各异质数据库。②局部模式：成分数据库概念模式，即本地数据库模型。③成分模式：从局部模式转换到联邦模式所用的一致性的数据模型，即联邦模式所采用的模型。④输出模式：成分模式的一部分，包含所有被输出到联邦模式的数据，过滤掉成分模式的私有数据。⑤联邦模式：总体模式，多个输出模式一体化集合。⑥外部模式：FDBS 用户及应用使用的模式。

注：完全实现联邦模式存在困难，更为可行的做法是建立**基于空间元数据的分布式结构实现元数据共享管理**。



- 4) **分布式空间数据库的优点总结**：①符合地理数据分布特点。②提高可靠性和可用性。某节点出故障，其他节点可继续使用，只是故障节点数据和软件不能使用。③实现局部自治的数据共享，不强调集中控制。④各节点对本身拥有的数据库有自治权、自主权，其他节点也可以共享这些本地数据。⑤硬件、软件和数据独立性。

4. **空间数据共享平台框架**：空间数据共享平台是指通过**网络平台**和**空间数据平台**建设，实现对多种类型数据的整合，为各种应用提供不同层次的技术平台支持与服务。

根据结构和功能的不同，可以将**空间数据共享平台**的总体框架分为三个层次：**共享基础平台**、**共享服务体系**和**共享应用体系**。

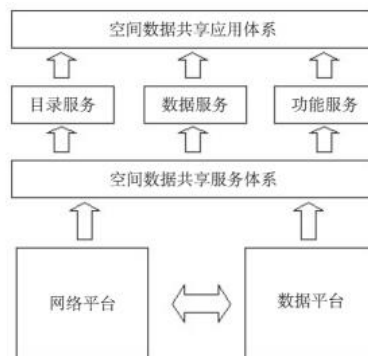


图2-20 空间数据共享框架（原图3-13）

- 1) **基础平台**包括基础网络平台和数据平台，是空间数据共享平台的基础与保障。
- 2) **空间数据共享服务**主要包括目录服务、数据服务和功能服务三个层次的内容。
 - a) **目录服务**：以空间元数据为核心实现目录查询与管理。
 - b) **数据服务**：空间元数据目录服务基础上提供空间数据服务，重点对空间数据提供浏览、查询和下载等常规服务。

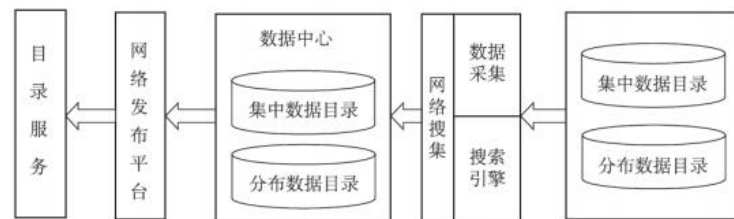


图2-21 目录服务框架原理图（原图3-14）

- 3) **共享应用体系**：共享技术平台功能体现，为用户或应用直接提供所需的共享数据、服务和各种应用功能。
 - a) **基础服务模块**：一系列符合相应标准和规范的基础性功能模块(如中间件)，为构建应用服务系统提供服务。
 - b) **空间数据论坛**：为满足空间数据应用、服务和研究需要，建立的科技论坛，为研究者和用户提供交流平台。
 - c) **空间数据智能服务**：空间数据共享平台的一个较高和较深层次应用。
5. **空间数据安全机制**：空间数据的安全主要包括**空间数据访问安全、传输安全及机密空间信息隐藏**等几个方面。
 - 1) **访问安全**：指网络 GIS 用于控制用户浏览和修改数据能力的方法，这些方法包括对数据逻辑视图浏览和用户获取空间数据的授权。
 - 2) **传输安全**：空间数据传输安全主要是指网络 GIS 用于控制空间数据网上传输不被非法截获、复制和修改的能力，保证空间数据的安全性与保密性。
 - ✧ **加密方法**：对称加密、非对称加密
 - ✧ **常用网络加密技术**：链路加密（中间节点明文加密，在线）、节点加密（中间节点安全模块密文加密，在线）、端-端加密（脱线加密）
 - 3) **机密信息隐藏**：信息隐藏是将一个消息(称为待隐消息或秘密消息)隐藏在另一个消息(称为遮掩消息或载体)中。
 - ✧ **原则**：既要将在机密空间信息隐藏起来，又不影响隐藏了机密信息后的空间信息的使用价值。

◇ 算法原理:

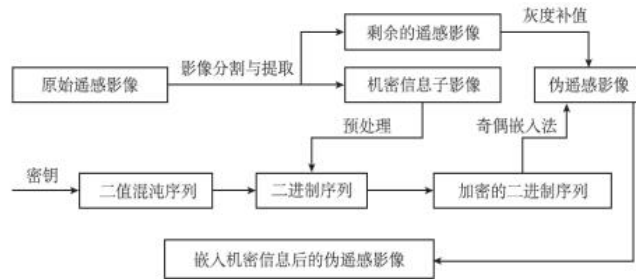


图2-23 影像机密信息隐藏算法框图 (原图3-23)

- 首先对原始影像进行机密信息分解、分析与综合，将机密地物从影像中识别出来，无论它是何形状，均可用其最小外接矩形选定，然后将包含机密地物的最小外接矩形从影像中提取出来，称为机密信息子影像；
- 对影像中提取机密信息后的“空白”影像块，根据其周围地物的地貌、形状与纹理特征，对其进行灰度补值，生成抹去机密信息但其余地物地貌特征均没有变化的伪影像；
- 对机密信息子影像进行压缩与二进制编码，并引入扩频通讯技术以及用密钥生成的二值混沌序列对机密信息进行加密；
- 采用奇偶嵌入法与 JPEG 标准量化表将机密信息以不可见的方式嵌入到伪影像中，即在伪影像的灰度补值影像块的空间域上内嵌机密信息，生成隐藏了机密信息的伪影像。

➤ 网络 GIS 通信协议及规范

- 通用协议与规范：TCP/IP、HTTP、HTML、XML
- 空间数据相关标准与规范：GML、SVG、KML、GeoJSON、GeoVRML
- TCP/IP：规定数据传输格式和传输方式，确保数据传输成功
 - TCP: Transmission Control Protocol, 传输控制协议
 - IP: Internet Protocol, 网际协议
- HTTP：超文本传输协议 (Hyper Text Transfer Protocol)，HTTP 提供了 WebGIS 运行的基本功能，采用请求/应答模型实现客户机与服务器的信息通信。
- HTML：超文本标记语言 (Hyper Text Markup Language)，HTML 是文本文件，代码集合，不是程序设计语言，定义的各种标识符由一些尖括号 “<”、“>”括起来，在浏览器端显示不同信息。

6. **XML：可扩展标记语言 (Extensible Markup Language)** 是 W3C 为适应 WWW 的需要，将 **SGML(Standard Generalized Markup Language, 标准通用标记语言)** 简化而成的标记语言，其功能比 HTML 更强大，不再是固定标记，**允许定义数量不限的标记**来描述文档中的数据，**允许嵌套的信息结构**，并提供了一种直接处理 Web 数据的通用方法。
7. **编码规范、标记语言的家谱表：**

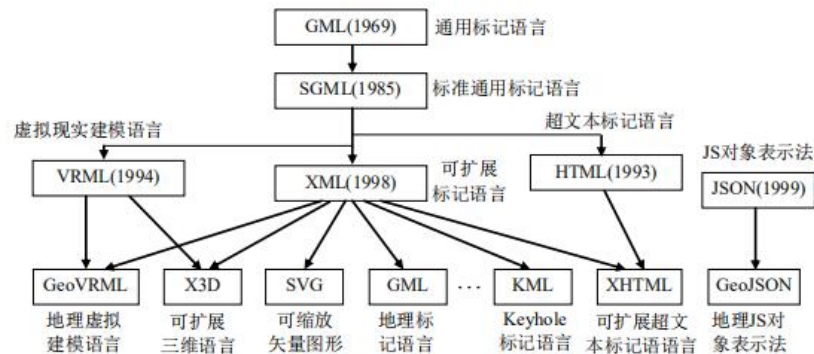
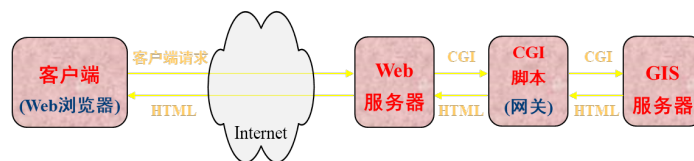


图2-24 标记语言家谱表 (新图)

8. **GML：地理标识语言 (Geography Markup Language)** OGC1999 年提出，是专门用于表示空间和属性数据的标记语言规范，是 XML 在地理空间信息领域的重要应用，特征由一些非空间属性和几何信息组成。
9. **SVG：可伸缩矢量图形 (Scalable Vector Graphics)** W3C 组织开发，利用 XML 描述二维矢量图形，由图形、影像和文字三个基本部分组成，三部分之间可以任意组合运用，可以保证图像显示质量不会因为缩放而失真或受损。
10. **KML：Keyhole 标记语言 (Keyhole Markup Language)**，Google 旗下 Keyhole 公司开发和维护 (2008 年被 OGC 宣布为开放地理信息编码标准，改由 OGC 维护和发展)，利用 XML 语法格式描述点、线、面、多边形和模型等。KML 很多方面与 SVG 类似，不同的是，SVG 中画布是二维的计算机屏幕，KML 中是地球表面。
11. **JSON：JavaScript Object Notation (JavaScript 对象表示法)**，一种轻量级数据交换格式，易于阅读、编写、机器解析和生成等。
12. **GeoJSON：使用 JSON 来编码地理空间数据**，使用 {key:value} 保存对象，表达地理特征、属性和范围，无严格的闭合标签，无需关注数据本身拓扑，格式简洁、数据利用率和传输效率高。
13. **GeoVRML：地理虚拟建模语言 (Virtual Reality Modeling Language)**，以 VRML 为基础来描述空间数据，利用标准 VRML 插件来浏览地理参考数据、地图。

➤ 网络 GIS 实现技术

1. **服务器端实现技术**：CGI 技术、Server API、动态网页技术
2. **客户端实现技术**：Plug-in 件技术、ActiveX 技术、Java Applet 技术
3. **CGI 技术（网关接口，Common Gateway Interface）**：使用户可以通过浏览器进行交互操作，并得到相应操作结果，可用 C、C++、Perl、Python 语言、Shell Script 及 Batch 等进行 CGI 编程。
 - 1) 通过 CGI 脚本将 GIS 服务器和 Web 服务器连接，**所有 GIS 操作和分析均在 GIS 服务器上完成。**
 - 2) GIS 服务器和 Web 服务器可置于 1 台主机，或多台机器
 - 3) **工作流程**：
 - a) Web 浏览器用户向 Web 服务器发出 GIS 相关功能请求。
 - b) Web 服务器接受请求，并通过 CGI 脚本将用户请求传送给 GIS 服务器。
 - c) GIS 服务器接受请求，进行数据处理，如：缩放、漫游、查询、分析等。
 - d) GIS 服务器将处理结果通过 CGI 脚本、Web 服务器返回给浏览器。



4. **Server API 模式**：服务器应用程序接口模式，依附于特定 Web 服务器，不像 CGI 可单独运行，它**运行于 Web 服务器进程中，启动后一直处于运行态，无需重新启动**，效率高于 CGI 程序，但程序可移植性差。
5. **动态网页：Active Page**，运行于 Web 服务器的页面，内嵌程序，服务器执行并将结果写入 HTML 文件流，再返回。

常见的动态网页技术：Microsoft ASP(Active Server Page)、Sun JSP(Java Server Page)、PHP(Hypertext Preprocessor)、Code Fusion。

◇ ASP 是微软公司推出的服务器端组件，能与 HTML 集成，易于创建，能自动编译和连接，客户端仅需一个浏览器，无其他特殊要求。

6. **Plug-in 技术**：Plug-in(插件)，扩展浏览器功能，使其支持空间数据处理是网络 GIS 的一种重要实现技术。Plug-in 可使 Web 浏览器支持特定格式的 GIS 数据处理，为 Web 浏览器与 GIS 服务程序之间通信提供条件。能直接处理来自服务器的矢量数据，但需

要先安装后使用。

工作流程：

- 1) Web 浏览器向 Web 服务器发出数据请求。
- 2) Web 服务器处理请求，将所需的 GIS 数据传给 Web 浏览器。
- 3) 客户端对接收的 GIS 数据类型进行分析和理解，如不需要 GIS Plug-in，则直接显示，如需要 GIS Plug-in 支持，则转往下一步。
- 4) 在浏览器中搜索相关的 GIS Plug-in，若有，则直接调用并显示 GIS 数据；若无，则从服务器上查找、下载并安装，加载到客户端以显示 GIS 数据。

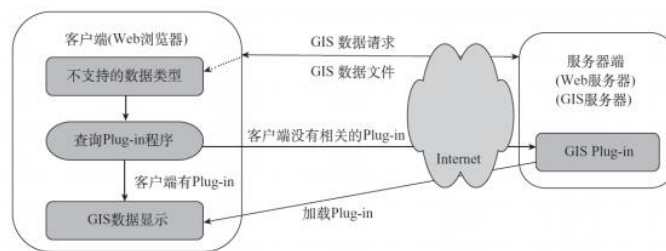


图2-27 基于Plug-in模式的WebGIS体系结构（原图5-3）

优点：①客户端处理能力较强，服务器与网络负荷较轻，对于 GIS 服务器而言，只需提供空间数据；②支持多源空间数据。

缺点：①平台相关性，插件方式受限于所用平台；②数据相关性。为显示和处理不同来源和格式的空间数据，需在浏览器上安装不同插件，且插件管理复杂化；③客户端功能有限。浏览器上完成的多是 GIS 的基本功能，对于较复杂的空间分析和智能处理操作有限

7. **ActiveX 技术：**微软提出的一种建立在 OLE 标准之上的规范和公共框架。与 Plug-in 一样，用于扩展 Web 浏览器功能。区别在于：Plug-in 与浏览器相关，ActiveX 能使用于任何支持 OLE 标准的程序或应用系统中。

工作流程：

- 1) Web 浏览器向 Web 服务器发出数据请求。
- 2) Web 服务器处理请求，配合 GIS 服务器将所要的 GIS 数据传送给 Web 浏览器。
- 3) 若客户机安装了 GIS ActiveX 控件，则无需再下载，否则需将 GIS ActiveX 控件下载并安装到浏览器上。
- 4) 浏览器利用 GIS ActiveX 控件对 GIS 数据进行相应的处理



图2-28 基于ActiveX的WebGIS体系结构（原图5-4）

优点：软件复用能力强。ActiveX 控件可以用多种语言实现。

缺点：①平台相关。不同的 GIS 平台须提供不同 ActiveX 控件。②兼容性差。ActiveX 是微软的一种规范，得到 IE 全面支持，**只能在 Windows 平台运行**，在其他浏览器中需定制控件才能运行。③预先下载。当浏览器上没有相应的 GIS ActiveX 时，**必须从服务器上先行下载，占用客户端磁盘空间**；④安全性低，**GIS ActiveX 能对磁盘进行读/写操作**。

8. **Java Applet**：let 技术运行在浏览器中的 Java 小程序，可视为 Java 插件。通过<applet>标签被嵌入在 HTML 文件中，其执行代码同时被下载到浏览器上，并由浏览器解释执行。

工作流程

- 1) Web 浏览器向 Web 服务器发出数据请求，访问服务器端 CGI 程序；
- 2) Web 服务器处理请求，由 CGI 程序将运算结果传送给 Web 浏览器。若客户机已安装 GIS Java Applet 控件，则无需下载，否则将 GIS Java Applet 控件自动下载并安装到浏览器上；
- 3) 浏览器利用 GIS Java Applet 控件对空间数据进行相应处理。

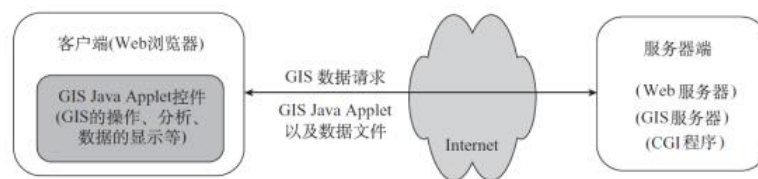


图2-29 基于Java Applet的WebGIS体系结构（原图5-5）

优点：①平台无关。可在 JVM 上解释执行，对浏览器和操作系统无特殊要求。动态运行。②Java Applet 在 WebGIS 运行时从服务器端下载，当 Java Applet 更新后，客户端总能及时更新

缺点：客户端负荷较重。计算集中于客户端，导致客户端负载较重，速度较慢。

9. **网络 GIS 实现技术的总结：**

类型	工作模式	案例	优点	缺陷
基于CGI的WEB GIS	CGI	IMS, ProServer	客户端很小；充分利用服务器的资源	JPEG和GIF是客户端操作的唯一形式；互联网和服务器的负担重，CGI的应用程序一般都是可执行程序。
基于Server API的WEB GIS	Server API	GeoBeans IMS	客户端很小；充分利用服务器的资源，动态连接库的形式存在。	JPEG和GIF是客户端操作的唯一形式；互联网和服务器的负担重。
基于Plug-in的WEB GIS	Plug-in	MapGuide	具有动态代码模块。比HTML更灵活，可直接操作GIS数据。	与平台和操作系统相关；不同的GIS数据需要不同的Plug-in支持；必须安装在客户机的硬盘上。
基于ActiveX的WEB GIS	ActiveX	GeoMedia Web Map	具有动态代码模块。通过OLE与其它程序、模块和互联网通讯。是一种通用的部件。	需要下载、安装，占有硬盘空间；与平台和操作系统相关；不同的GIS数据需要不同的ActiveX控件支持。
基于Java Applet的WEB GIS	Java Applet	ActiveMap, GeoBeans	在支持Java的互联网浏览器上运行，与平台和操作系统无关；完成GIS数据解释和GIS分析功能。	对于处理较大的GIS分析任务的能力有限；GIS数据的保存、分析结果的存储和网络资源的使用能力有限。

表2-4 WebGIS技术比较（原表5-1）

技术类别		通用网 关接口	服务器应用 程序接口	动态网 页技术	插件 技术	Java Applet 技术	ActiveX 技术
性能指标	客户端	很好	很好	很好	好	好	好
	服务器端	差—好	好	好	好	很好	很好
	网络负荷	重	较重	较重	较轻	较轻	较轻
	综合运行能力	一般	好	好	好	好—很好	好—很好
交互能力	用户界面	差	好	好	好	很好	很好
	功能支持	一般	好	好	好	很好	很好
	本地数据支持	否	否	否	是	否	是
可移植性	整个系统	差	很好	差	差	好	一般
安全性	整个系统	很好	很好	很好	一般	好	一般

第三章 · 空间数据网络存储

➤ 数据存储技术

1. **数据存储**：以某种格式将数据记录在计算机内部或外部存储介质上。
2. **数据存储技术的发展**：
 - 1) **磁盘存储技术**——长盛不衰，仍有突破
存储密度从 1Tbits/in² 增加到 5Tbits/in²；存储容量将达到 20TB 以上
 - 2) **光存储技术**——与磁存储技术并行发展
超高密度光存储设备可提供高达 1000TB 的存储容量
 - 3) **可生存存储技术**——热点研究领域：永久可用、永久安全和平稳降级
 - 4) **基于微电子机械系统的存储**：性能更高、容错能力更强、廉价
 - 5) **DNA 存储技术**：每个脱氧核苷酸含 4 种碱基：腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)、胞嘧啶(C)。他们的排列顺序决定生物多样性。若用 0、1、2、3 各代表一个碱基，可以组成一个四进制存储模式。
 - 6) **网络存储技术**：得益于各项存储技术的进步，云存储、SAN、NAS 等网络存储技术将在网络结构、存储机理等方面继续完善。

3. **数据存储技术的分类——按连接方式分类**（即按伺服主机与存储设备间的数据存取信道进行分类）：

- 1) **直接连接存储(Direct Attached Storage, DAS)**：存储设备与服务器直接连接的存储方式。
- 2) **附网存储(Network Attached Storage, NAS)**：将存储设备直接附着在基于 TCP/IP 协议的现有以太网上的一种存储技术；以文件为单位。
- 3) **存储区域网(Storage Area Network, SAN)**：一个使用光纤通道连接的高速专用存储子网，不占用服务器的资源，也不占用服务器所属的通信网络的带宽，通过 SCSI 命令服务器之间通信；以数据块为单位。

4. **数据存储技术的分类——按存储系统联机方式分类**（即按存储介质的特性与数据存取的效率进行分类）

- 1) **在线存储(On-line Store) ——工作级存储**：采用高速数据存储设备，常常是高端 SCSI 存储设备，以便能随时读写数据，满足对数据访问的速度要求。
- 2) **离线存储(Off-line Store) ——备份级存储**：是一种无需时刻与计算机保持联系的存储设备，主要用于对在线存储数据的备份，以防范可能发生的数据灾难。
- 3) **近线存储(Near-line Store)**：是介于离线存储和在线存储之间的一种技术；存储非实时需求的数据或备份重要数据，如光盘库。

相关技术：数据迁移（也称为 HSM，分级存储管理）：将大容量离线存储设备作为磁盘等在线存储设备的下级设备，将常用的数据按一定策略自动地从磁盘迁移到离线设备上，需要的时候再将这些数据自动从离线设备调回到磁盘上。【实质是将很少使用或不用的数据移到大容量离线存储设备的存档过程，是离线与在线存储技术的融合】

5. **数据存储技术的分类——按接口协议标准分类**

- 1) **小型计算机系统接口 (Small Computer System Interface, SCSI)**：一种高性能计算机外部设备接口
- 2) **光纤连接**：利用光纤通道技术来连接存储设备
- 3) **第三代输入输出总线 (3GIO)**：支持各种全新的数据交换技术，用在 PC、便携式计算机、服务器等设备上，为网络数据交换提供更高性能，现在叫 PCI-Express 总线和接口标准，由 PCI-SIG 开发并由 Intel 率先采用。

- 4) **InfiniBand**: 一种服务器 I/O 技术规范新的网络互连技术, 采用基于包交换的高速交换网络技术, 实现服务器内部的互连、服务器之间互连、集群系统互连以及存储系统互连。
6. **磁盘阵列技术 (RAID, Redundant Array of Inexpensive Drives)**:
- 1) 是一种提升数据**存取能力和存取效率**的存储技术;
 - 2) 将**多个存储设备按照一定的方式集成起来**, 组成一个**逻辑上单一**的大容量存储设备。数据分布在多个磁盘中, 所以在进行数据输入/输出时, 多个磁盘能同时工作, **并行传输**数据, 从而提高整个系统的数据传输速度和吞吐量
 - 3) **RAID 技术包括两个方面**: 数据的安全性与磁盘的执行效率。
 - 4) **RAID 的特点**: ①存储容量成倍增大; ②数据存取速度极大提高; ③容错能力显著改善。
 - 5) 磁盘阵列可分为软件**磁盘阵列 (software RAID)**和**硬件磁盘阵列 (hardware RAID)**两种, 硬件磁盘阵列又分为**内置式**和**外置式**两种。
 - 6) **RAID 几种代表技术**:
 - a) **RAID0**: 采用**数据分条**技术(Data Stripping Array without Parity), 即把数据分成若干个**大小相等**的小块(Block), 经由 RAID 控制器**平均分配**到阵列的磁盘中, 因此又称为“Stripping”(即将数据条带化)。
 - 优点: 成本低、读写性能和空间利用率高; 缺点: 安全性和可靠性较低, **没有任何容错措施, 不提供数据的冗余处理。**
 - b) **RAID1**: 采用**磁盘镜像技术**(Disk Mirroring), 即数据完全冗余。其原理是在两个磁盘之间建立完全的镜像。
 - 优点: 其他 RAID 为数据提供了更高的安全性和可靠性;
 - 缺点: 磁盘空间利用率低(不到 50%), 存储成本高
 - c) **RAID0+1**, 也称为 RAID10, 是 RAID0 和 RAID1 技术的组合, 以充分利用这两种技术的优点, 达到数据存储和传输既高速又安全的目的。【可以看成是两个 RAID0 互为镜像】
 - d) **RAID3**: 采用**单盘容错并行传输技术**(Parallel Disk Array)。其工作原理是采用 Stripping 技术把数据分成**多个“块”**, 并按照某种容错算法对数据进行**位校验**。

- 该种阵列有 $n+1$ 个磁盘组成，其中 n 个磁盘作为数据阵列，第 $n+1$ 个作为容错盘，存储容错信息；当 $n+1$ 个硬盘中有一个硬盘出现故障时，通过容错盘和其他磁盘的联合作用可以恢复原始数据；然而，若有多于一个磁盘的数据产生错误，则无法进行纠正。
- e) **RAID5**: 采用旋转奇偶校验独立存取技术，是一种存储性能、数据安全和存储成本兼顾的存储解决方案，目前应用最为广泛，其原理是：将各块独立磁盘进行条带化分割，相同的条带区进行奇偶校验，校验数据平均分布在每块磁盘上。
- 不对存储的数据进行备份，也没有固定的校验盘，而是按某种规则把数据和相应的奇偶校验信息均匀地分布存储到各磁盘上，每块磁盘上既有数据信息也有校验信息。

表3-1 常用RAID级别的特点比较(原表4-1)

RAID 级别 比较类别	RAID0	RAID1	RAID3	RAID5
容错性	无	有	有	有
冗余类型	无	有(完全冗余)	奇偶校验	奇偶校验
热备份选择	无	有	有	有
硬盘要求	一个或多个	偶数个	至少三个	至少三个
有效硬盘容量	全部硬盘容量	硬盘容量 50%	硬盘容量 $(n-1)/n$	硬盘容量 $(n-1)/n$
存储方式	数据分块循环存到各硬盘	数据分块循环存到各硬盘	数据以字节交叉方式存于 n 个硬盘	数据以块交叉方式轮流存入 n 个硬盘

表3-2 常用RAID级别的性能比较(原表4-2)

RAID 级别 比较类别	RAID0	RAID1	RAID3	RAID5
读取数据	容易处理多个同时读取	较快,其中的任何一个硬盘都有数据	正常速度,与一个的速度一样	快
写入数据	容易处理多个同时写入	较慢,需要重复写入多个硬盘	较慢, Parity 编码的运算包含从其他硬盘内读取与写入 Parity 编码所需要的时间	较慢, Parity 的计算包含读与写
备份功能	无	安全性最高	很好	很好
空间利用率(费用)	非常合理,硬盘空间完全利用	硬盘使用率不到 50%。	合理,硬盘空间被充分利用	合理,硬盘空间被充分利用

7. 数据存储接口协议和标准——SCSI 技术:

- 1) 原是为小型机和工作站研制的一种传输速度高的接口技术，现在已被广泛使用于普通设备上，如硬盘、光驱、ZIP、MO、扫描仪、磁带机、JAZ、打印机、光盘刻录机等设备

2) SCSI 的类型:

- a) **SCSI-1**: 原始的版本, 传输时的频率异步时为 3MHz, 同步时为 5 MHz, 几乎被淘汰了。
- b) **SCSI-2, 又称 Fast SCSI**: 同步传输时的频率使数据传输率从 5Mbps 提高为 10Mbps, 支持 8 位并行数据传输, 可连接 7 个外设; Wide SCSI, 支持 16 位并行数据传输, 数传率也提高到了 20Mbps, 可连接 16 个外设。
- c) **SCSI-3, 又称 Ultra SCSI**: SCSI-3 将同步传输时钟频率提高到 20MHz, 数传率达 20Mbps。若使用 16 位传输的 Wide 模式, 数传率可提高至 40 Mbps。

3) SCSI 的优点: ①适应面广, SCSI 可支持多个设备; ②多任务、高性能, SCSI 允许在对一个设备传输数据的同时, 另一个设备对其进行数据查找; ③CPU 占用率少由于 SCSI 卡本身带有 CPU, 可处理一切 SCSI 设备的事务, 在工作时, 主机 CPU 只要向 SCSI 卡发出工作指令, SCSI 卡便可独自进行工作, 工作结束后将结果返回给主机 CPU; ④智能化, SCSI 卡可对 CPU 指令自动进行排队; ⑤带宽高, 最快的 SCSI 总线有 160Mbps 的带宽(需要 64 位 66 MHz 的 PCI 插槽); ⑥可用于外置或内置。

4) SCSI 的缺点: 在同样条件下, 因为 SCSI 硬盘的控制指令比 IDE 硬盘的控制指令复杂, 所以 SCSI 硬盘内部传输速度要比 IDE 硬盘慢。

5) SCSI 与 IDE 的区别:

IDE(Integrated Drive Electronics, 集成驱动器电子接口), 也称为 AT-Bus 或 ATA 接口, 同样也是极为常用的一种硬盘接口。

- a) IDE 的工作方式需要 CPU 的全程参与, CPU 读写数据的时候不能再进行其他操作, 从而导致系统性能下降。
- b) 目前的 IDE 接口针对该问题做了很大改进, 已经可以使用 DMA 模式而非 PIO 模式来读写数据, 数据交换由 DMA 通道负责, 对 CPU 的占用大为减小。

8. 数据存储接口协议和标准——光纤通道技术:

- 1) 光纤通道技术(Fibre Channel, FC) 的设计集成了通道技术和网络技术的优点, 可应用于高速、高可靠和可扩展的主干网络。
- 2) 光纤通道的传输距离很大程度上取决于所使用的**线缆类型 (光纤线缆、铜缆) 和**

传输率。传输率越高，线缆的直径越大(对于光纤线缆)，传输的距离就越远。

- 3) 光线通道技术的结构层次模型：逻辑上光线通道标准定义了功能类似的五个模块化层次。



图3-2 光纤通道结构层次模型(原图4-2)

- 4) 光纤通道有**点对点**、**仲裁环路**和**交换 fabric** 三种基本形式的**拓扑结构**。这些拓扑结构能够组合起来构成满足特殊需要的 SAN 系统。
- 点对点拓扑结构**：应用于两个设备之间，无需解析地址，所有的帧(光纤通道数据包)全部发给另一个设备。
 - 仲裁环拓扑结构**：也称光纤通道仲裁环路(FC-AL)，是以环状形式连接所有设备的拓扑结构。在这种结构中，一台设备从上游设备接收光纤信号，并将其发送给下游设备，其中环路上的设备通过一个 8 位的 AL_PA(仲裁环路物理地址)来标识。
 - 交换 fabric 拓扑结构**：是一种允许连接大量设备的拓扑结构，易于扩展，可热插拔设备，它允许交换机使用硬件电路在节点间路由数据，从而更有效地使用带宽。
- 5) **服务等级**：为描述不同端口在传输数据时采用的交互机制，光纤通道定义了多个服务等级，分别为等级 1~4 和等级 F。
- 等级 1**：是一个面向连接的服务，建立了发送者和接收者之间的专用连接运行在光纤上的虚拟端对端的链路，保证了两个 N/NL 端口之间的连接。
 - 等级 2**：是一个面向无连接的服务，提供两个端口之间无连接的通道。等级 2 的帧同时使用缓存到缓存和端对端的流控制，并允许设备共享所有可用的带宽。
 - 等级 3**：是多元的无连接的数据报服务，为光纤通道网络中最常用的服务等级。数据报服务是指不包含传输确认的通信。

d) **等级4**：是面向连接的部分带宽服务，中将带宽划分为独立的虚拟电路，以安全分配带宽资源。

e) **等级F**：用于 fabric 中的内部控制和协调

9. 数据存储接口协议和标准——iSCSI 技术：

iSCSI(Internet Small Computer System Interface, 互连网小型计算机系统接口) 是一个供硬件设备使用的、可在 IP 协议上层运行的 SCSI 指令集，是一种开放的基于 IP 协议的工业技术标准。【基于 IP 存储理论的新型存储技术，该技术是将存储行业广泛应用的 **SCSI 接口技术与 IP 网络技术相结合**，使得可在 IP 网络上组建 SAN。】

1) **iSCSI 的工作流程**：iSCSI 协议规定了在网络上封包和解包的过程。在网络的一端，数据被封装成包括 IP 头、iSCSI 识别包和 SCSI 数据在内的三部分内容，传输到网络另一端时，三部分内容分别被顺序解开。

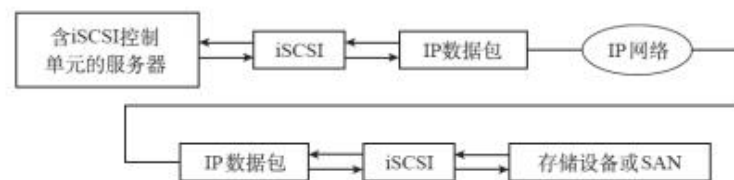


图3-3 iSCSI的工作流程(原图4-3)

2) **iSCSI 的安全性**：iSCSI 采用 IP 网络技术作为底层传输技术，与现有的通信网络是完全兼容的，因而 iSCSI 具有 IP 网络类似的常见安全性问题，针对各种安全风险，iSCSI 一般采用两种安全措施：①认证：在目标机器与发起者之间进行身份认证；②加密：对传输的 TCP/IP 数据包进行加密保护。

➤ 网络存储分类

1. **直接连接存储 (DAS) /基于服务器附属存储 (SAS, Server Attached Storage)**：“**以服务器为中心**”的存储结构，以并行 SCSI 总线连接至某一特定的主机，存储设备只能被该主机直接访问和控制，其它主机需要访问存储设备中的数据时，必须经该服务器的存储和转发。

✧ **效率低下，实时性差，并且存储 I/O、网络 I/O 以及 CPU 和内存均较易成为系统的瓶颈**

✧ **DAS 数据存取步骤**

1) 请求命令经过网络发至服务器；

- 2) 服务器查询缓冲区，若数据在缓冲区中则经网络将其直接发给客户机；否则将请求翻译成本地数据访问命令并发向与服务器相连的存储设备；
 - 3) 存储设备接收到命令后将数据复制到服务器的系统缓冲区；
 - 4) 系统缓冲区将数据复制到网络适配器(网卡)的数据缓冲区中；
 - 5) 最后通过网络将数据发给客户端；
2. 附网存储 (NAS, Network Attached Storage): 将分布、独立的数据整合为大型集中化管理的数据中心，以便于不同主机和应用服务器对其进行访问的技术。
- ✧ 主要特征是把存储设备和网络接口集成在一起，直接通过网络存取数据
 - ✧ NAS 设备的硬件结构由**控制器和存储子系统**两部分构成。

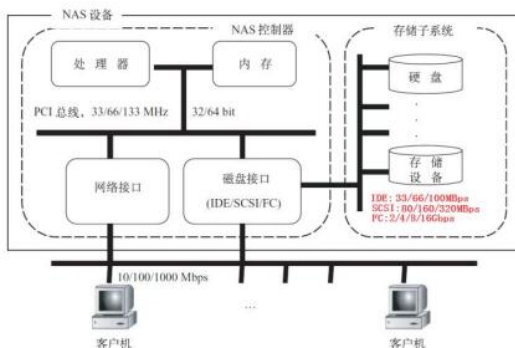


图3-6 NAS服务器的基本硬件结构(原图4-6, 红色部分有更改)

- ✧ 为实现较高的稳定性和 I/O 吞吐率，满足数据共享、数据备份、安全配置、设备管理等要求，NAS 系统软件结构设计为五个模块：**操作系统、卷管理、文件系统、网络文件共享和 Web 管理。**

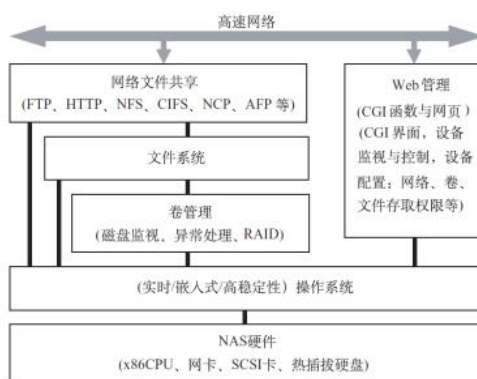


图3-7 NAS软件系统结构(原图4-7)

- ✧ NAS 完全以数据为中心，是文件级存储方法，通常用于文件服务。由工作站和服务端通过**网络协议**(如 TCP/IP)和**应用程序**(如 NFS 或者通用 Internet 文件系统 CIFS)进行访问。NAS 实际关注的是文件服务而不是实际文件系统的执行情况。

3. **存储区域网络 (SAN, Storage Area Network)**: 是一种新的以**数据存储为中心**的网络存储体系结构, 采用可伸缩的网络拓扑结构, 通过光通道直接连接方式为 SAN 内部任意节点提供多路可选择的数据交换, 并且将数据存储管理集中在相对独立的存储区域网内。

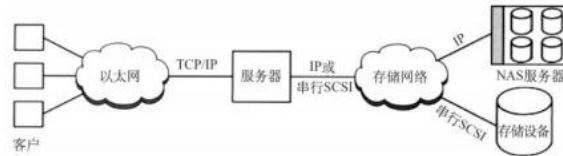


图3-8 SAN拓扑结构示例(原图4-8)

- ✧ SAN 系统由**光纤通道硬件产品**和**SAN 管理软件**构成。光纤通道硬件产品是构成 SAN 系统的物质基础, 提供服务器间和存储设备间的高速连接; 而 SAN 管理软件则相当于 SAN 系统中的操作系统, 决定着 SAN 系统的功能和性能。
- ✧ SAN 有如下图中所示的三种构造方法, 即**基于交换机的交换式 SAN**、**基于集线器的共享式 SAN**和**以交换机为主干的混合式 SAN**。

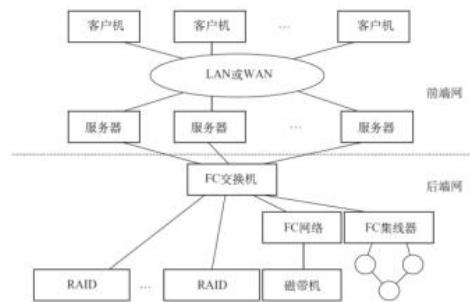


图3-9 SAN的构造方法(原图4-9)

- ✧ SAN 具有**可靠性**、**高性能**、**动态配置存储结构**三个主要特点。
4. **SAN 相对于 DAS 和 NAS 的优势:**
- 1) 采用**网络结构**, 具有**无限的扩展能力**, 用户可以根据实际需求自由增加存储设备, 以保证系统的整体性能。
 - 2) **通信协议与现有的应用系统完全一致**, 适应面广。在主机、存储设备之间使用 SCSI 协议进行通信, 在主机与主机之间采用 IP 协议通信。
 - 3) 采用**光纤通道技术**, 具有更高的连接速度和处理能力。
 - 4) 能够将所有客户机、服务器、存储设备、交换机、网络与存储管理工具等多种软硬件系统构成的共享存储池连接起来, 提供高速的共享资源访问服务。
 - 5) 集中存储管理服务模式有助于降低企业的数据存储开销。

- 6) 承担以前服务器的备份与恢复、数据迁移等任务,极大地提高了服务器性能。
- 7) 具有较好的数据完整性、可用性和可靠性

➤ 网络存储模式

1. **存储虚拟化技术**: 对存储硬件资源进行的抽象化表现,屏蔽存储子系统各物理存储设备的区别和复杂性,面向用户提供统一的存储资源和服务。

✧ 存储虚拟化可以在不同的层面上进行: ①主机/服务器; ②存储网络(交换机或存储专用设备); ③存储子系统(智能阵列控制器)

- a) **基于主机/服务器的虚拟化** (理解: 将翻译器安装在演讲者身上): 通常由逻辑卷管理软件(Logical Volume Manager)在主机/服务器完成, 经过虚拟化的存储空间可以跨越多个异构的磁盘阵列。
- b) **基于存储设备的虚拟化** (理解: 将翻译器安装在听众身上): 这种层次的存储虚拟化通过智能阵列控制器完成, 将阵列上的存储容量划分为多个逻辑单元, 供不同的主机系统访问。
- c) **基于存储网络的虚拟化** (理解: 会场中的同传设备): 基于网络的存储虚拟化, 通过使用专用的虚拟化引擎来实现, 适合于开放的存储网络, 即 Open SAN, 它独立于主机, 同时也独立于存储设备。所谓的虚拟化引擎(SAN Appliance)是一种用来完成虚拟化工作的专用存储管理服务器。

2. **网络存储集成式技术**:

NAS 和 SAN 是两种**互为补充**的存储技术, 分别提供对不同类型数据的访问。NAS 提供**文件级数据访问**功能, 而 SAN 则主要提供**海量、面向数据块的数据访问与传输能力**。SAN 在数据块传输和扩展性方面表现优秀, 并能够有效地管理设备。与 SAN 相比, NAS 支持多台对等客户机之间的文件共享。应该说, SAN 和 NAS 均基于开放的、业界标准的网络协议。SAN 与 NAS 在技术层面的互补性表现如下:

- 1) NAS 设备逐渐采用 SAN 来解决与存储扩展和备份恢复相关的问题, 这将使得 NAS 和 SAN 之间的许多原有差别逐步消失;
- 2) 将 NAS 与 SAN 连接起来的“存储管理+应用”专用服务器能够使 SAN 像访问网内存储设备一样访问 NAS;
- 3) 将 SAN 与 NAS 服务器连接, NAS 服务器可以代替 SAN 中的文件服务器, 同时也可提供文件共享服务。

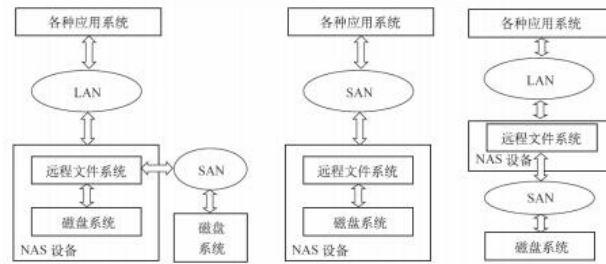


图3-10 NAS与SAN的集成(原图4-10)

➤ 云存储

1. **云存储技术(Cloud Storage):** 是以**云计算技术为基础**，结合多种存储设备和存储模式起来的一种网络存储技术，其实质是一个**以数据存储和管理为核心的云计算技术**。

特点：①按需服务；②弹性伸缩；③成本低廉；④资源利用率高；⑤安全可靠

2. **云存储的分类（根据存储服务对象分类）：**

- 1) **公有云存储(Public Cloud Storage):** 第三方云服务商提供，廉价共享存储资源，通过 Internet 访问。
- 2) **私有云存储(Private Cloud Storage):** 供单一企业内部各用户共享存储资源，高安全性云存储模式。
 - a) **自建模式：**运行于 Intranet、物理隔离，需软硬件设施、专门运维人员，投入高
 - b) **托管模式：**存储服务商提供托管服务，软硬件部署在安全的主机托管场所；或从其公有云中划分出一定的空间为企业提供私有云服务
- 3) **混合云存储(Hybrid Cloud Storage):** 数据被分存于私有云和公有云中。**私密性高、更新和访问频次低**的数据更适合存入私有云/本地存储系统中，而**访问和更新频繁**的数据则部署在公有云存储中。



图3-14 混合云存储示意图(新图)

3. 云存储的分类（按存储技术和访问方式分类）:

- 1) **块级云存储**: 数据块级存储服务, 直接对二进制数据操作。
- 2) **文件级云存储**: 基本操作对象是文件和目录, 存储位置由文件系统确定, 文件系统负责将文件目录和文件内容映射到底层的磁盘存储区域中。
- 3) **对象级云存储**: 对象存储结合了文件级和块级存储两者优点, 以数据对象为基本存储单元, 单元内含文件数据和一组元数据信息。舍弃了目录层次结构, 改而采用容器和对象的两层扁平结构进行管理。每个对象唯一地从属于一个容器, 实现基于容器的对象分组。

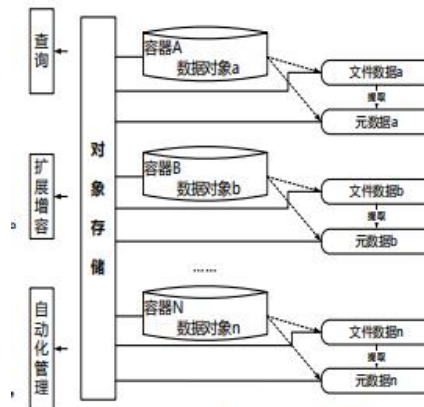


图3-15 对象存储结构图(新图)

- ✧ 在开源软件领域, OpenStack 平台的 Swift 和 Ceph 均为对象存储系统的代表。其中, OpenStack Swift 是一个开源项目, 面向大规模非结构化数据的存储管理, Swift 支持对数据的高效存储、快速检索与更新, 基于**一致性哈希技术**将存储对象均匀散列到虚拟的存储节点上。

4. 云存储技术的架构: 存储资源层、基础管理层、接口层和访问层。

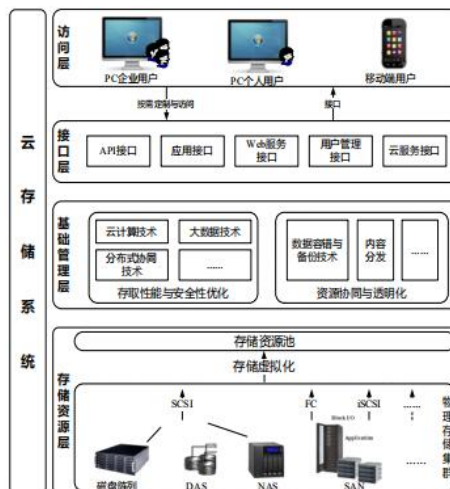


图3-16 通用云存储系统架构图(新图)

➤ 网络 GIS 数据存储实例

1. **商用网络化存储解决方案——IBM SAN 方案：**IBM SAN 的基本内容包括**连接、管理、使用和服务**，它将**光纤通道集线器、交换机和网关**等硬件与软件管理功能结合为一体，各种设备和软件不论是否出自 IBM 公司都可以密切配合，随时随地实现数据的**存储、访问、共享和保护**。
2. **商用网络化存储解决方案——IBM 云存储方案：**该方案提供了以全闪存阵列存储为基础、软件定义存储为架构、基于混合云平台的可扩展的、安全的大规模云端数据存储和管理技术，通过云平台向用户提供所需的连接和控制，以便在云上发掘、利用和扩大数据的业务价值，实现 IBM 提出的包括“安全、多云、开放和 AI Ready”四个层面的现代基础架构理念(Modern Infrastructure)
 - 1) **软件定义存储：**
 - a) IBM Spectrum Virtualize: IBM 提供的一种虚拟化软件层。
 - b) IBM Spectrum Accelerate: IBM 的软件定义块存储技术。
 - c) IBM Spectrum NAS: IBM 的软件定义文件存储技术，旨在简化 NAS 管理，提高敏捷性。
 - d) IBM Spectrum Scale: IBM 的软件定义高级非结构化数据存储与管理技术，基于主动文件管理(Active File Management, AFM)的分布式磁盘高速缓存技术
 - 2) **全闪存存储：**IBM 闪存的核心是**FlashCore 技术**，主要依靠提供基于硬件的数据路径，产生极低的存储响应延迟。
 - a) 主要特点：①端到端的完整架构设计；②提供硬件级别数据路径；③分布式 CPU 独立处理路径；④更低的延迟度。

第四章 · 移动 GIS 技术

➤ 移动 GIS 的组成

1. **互联网+：**是指在创新 2.0（信息时代、知识社会的创新形态）推动下由互联网发展的新业态，简单的说就是“互联网+传统行业”。
2. **移动地理信息服务：**是一种空间信息服务和应用模式，是移动通信技术与空间信息技术相结合的产物。
3. **移动 GIS** 是指将 **GIS、GNSS 和移动互联网一体化**的技术和系统，具有信息处理智能

化、信息服务个性化、信息来源多样化、位置服务实时化等优点。

4. **移动 GIS 的组成**：移动 GIS 主要由 **Internet**、**无线通信网络**、**移动终端设备**、**地理应用服务器及空间数据库**组成。

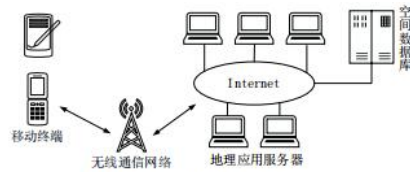


图4-2 移动GIS组成结构

- 1) **Internet**：网络传输、存储和计算环境，连接着地理应用服务器、空间数据库、各种外部设备等，并经由无线通信网络与移动终端设备连接。
 - 2) **无线通信网络与定位系统**：
 - a) **GNSS**：用以定位移动设备。GPS、Glonass、Galileo、BD 是四种主要的全球卫星导航定位系统。
 - b) 基于蜂窝通信系统的 GSM、GPRS、CDMA、WCDMA、CDMA2000、TD-SCDMA、TD-LTE、FDD-LTE 等
 - 3) **移动终端设备**：移动 GIS 的客户端(移动终端)设备是一种便携式、低功耗、适合于地理应用，并且可以用来快速、精确定位和地理识别的智能设备。包括便携式计算机、掌上电脑、平板电脑、智能手机等。
 - 4) **地理应用服务器**：是系统的 GIS **引擎**，位于固定场所，为移动 GIS 用户提供大范围的地理服务，以及潜在的空间分析及查询操作服务。
5. **移动定位技术**：
- 1) **卫星定位**：GPS、GLONASS、GALILEO、BD
 - 【**GNSS：全球导航卫星系统 (Global Navigation Satellite System)**，泛指所有的卫星导航系统，包括全球的、区域的和增强的卫星导航系统】
 - 2) **手机定位**：高通公司 GPSone，利用 GPS 和手机定位进行混合定位。
 - 3) **惯性导航系统**：由惯性测量、控制显示、状态选择装置、导航计算机和电源等组成，通过测量运载器的加速度，自动进行积分运算 获得运载器瞬时速度和位置。
 - ✧ 缺点：误差随时间迅速积累
 - 4) **射频识别定位**：使用多种频段通信，完成电子标签识别和读写，可用于停车场、滑雪场等，由用户自己布置在特定区域，类似的有红外、超声波、蓝牙和超宽带等室内无线定位技术。

- 5) **室内定位技术 (室内导航的关键技术):** Wi-Fi、蓝牙、RFID (射频识别系统, 通过射频无线信号自动识别目标对象, 并获取相关数据, 由电子标签、读写器和计算机网络构成)、超宽带无线电(UWB)、航迹推算等计算人和物体在室内位置。

☆**室内导航:** 用户在建筑物内用精确的定位功能确定位置并找到想去的地方

6. **移动 GIS 特点:** ①移动性; ②频繁断接性; ③弱可靠性; ④非对称性; ⑤资源有限性; ⑥空间位置依赖性。

➤ 移动空间数据管理

1. **移动空间数据**是用来描述空间位置、空间关系、空间方向的载体, 受限于存储空间、计算能力、电池容量、网络带宽。
2. **移动空间数据管理与传输的两种解决办法:** ①优化存储格式、压缩编码、建立高效空间索引; ②通过金字塔和渐进式传输技术。
3. **移动空间数据的移动端数据组织:** 移动空间数据视为**移动客户端空间数据**, 主要有**文件存储**和**嵌入式数据库存储**两种存储方式。

- 1) **移动端—文件存储:** 调用相关接口加载缓存在移动端的 tpk 和 vtpk 地图底图文件, 也可加载 ArcGIS Desktop 支持的常见空间数据文件。

✧ tpk、vtpk 文件是 ArcGIS 推出的将地图底图切片文件打包形成离线地图包的标准文件格式, 通过将底图切片进行封装有利于底图数据的传输、下载和共享等网络操作

- 2) **移动端—嵌入式数据库:** 应用于嵌入式系统的数据库, 除了能够实现对结构化数据的管理, 还针对移动终端设备特点进行了体系结构优化, 具有轻量级、独立、无服务器组件、占用内存少等优点。

常见的嵌入式数据库包括: SQLite、Berkeley DB、LevelDB、UnQLite。

移动 GIS 中, SQLite 常与 SpatiaLite 配合使用, SpatiaLite 是扩展 SQLite 内核的开源库, SpatiaLite 使得 SQLite 支持**高效的空間查询功能**。

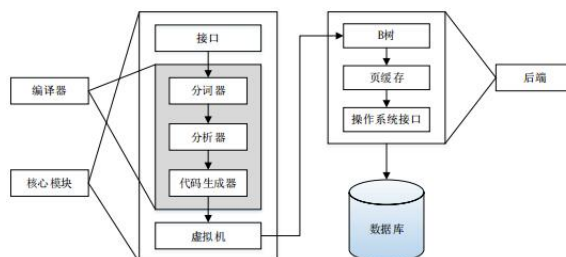


图4-3 SQLite基本架构

4. 移动空间数据的服务器端数据组织：ArcGIS for Server 发布的底图切片方案有紧凑型(Compact)、松散型(Exploded)两种模式

- 1) 紧凑型缓存的底图切片为紧凑型切片，扩展名为*.bundle 和*.bundlx，多在 ArcGIS 的服务器端使用。
- 2) 松散型缓存的底图切片为松散切片，扩展名多为.PNG，可以在非 ArcGIS 服务器中使用。

5. 移动空间数据压缩：

- 1) 矢量数据压缩：移动 GIS 中矢量数据压缩一般为有损压缩。主要有垂距限值法、角度限值法、光栏法、道格拉斯—普克(Douglas-Peucker)算法、基于小波的压缩方法。
 - a) 垂距限值法与角度限值法的基本原理是：设定垂距阈值 L (垂距限值)或角度阈值(角度限值)，从曲线的一端开始依次选择三个点 A、B、C，计算中间点 B 到直线 AC 的距离 d 或者直线 AB 与 AC 的夹角，如果大于阈值，则保留 B 点，否则舍弃。继续取下三个点计算，直到除端点外所有点计算完成为止。

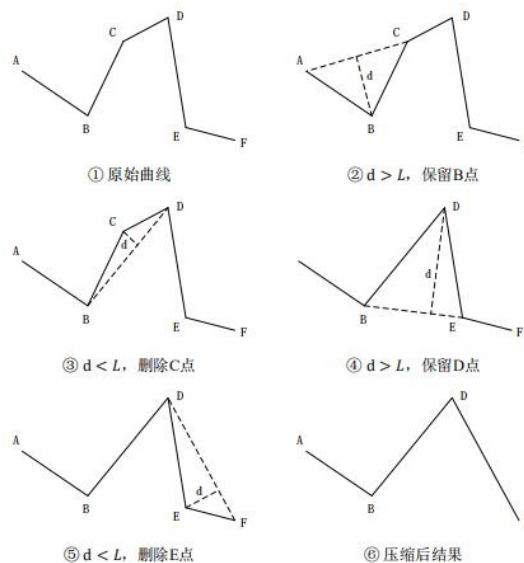


图4-4 垂距限值法示例

2) 栅格数据压缩：链式编码、游程编码、块码和四叉树编码

- a) 链式编码：按照顺时针方向依次定义方向：东为 1、东南为 2、南为 3、西南为 4、西为 5、西北为 6、北为 7、东北为 8。例如：原点坐标为(2,2)，则类别 F 区域的多边形要素按顺时针方向的链式编码为：

(2,2), 1, 1, 1, 3, 2, 4, 5, 6, 6, 7



图4-6 链式编码示意图

图4-7 栅格编码

- b) **游程编码**: 游程编码的基本思路是针对一幅栅格图像, 按照行(列)来记录栅格属性值和重复的像元个数, 也可以记录栅格属性值和游程的起始或终止行(列)号。假设某栅格直接编码如图 4-7 所示, 沿行方向进行游程编码的结果为: (0,2)(2,3)(5,2)(3,1); (0,3)(2,2)(5,2)(3,1); (2,2)(1,4)(3,2); (2,2)(1,2)(3,4); (4,4)(1,4); (4,4)(1,4); (5,2)(4,2)(1,4); (5,2)(4,2)(1,4)。(5,2)(4,2)(1,4)。

6. **移动空间数据传输——渐进式传输**: 从小比例尺概略表达到大比例尺详细表达的过程。当服务器端收到请求后, 首先向客户端快速传输分辨率低的轮廓数据, 然后根据用户需求逐步传输分辨率高的数据。

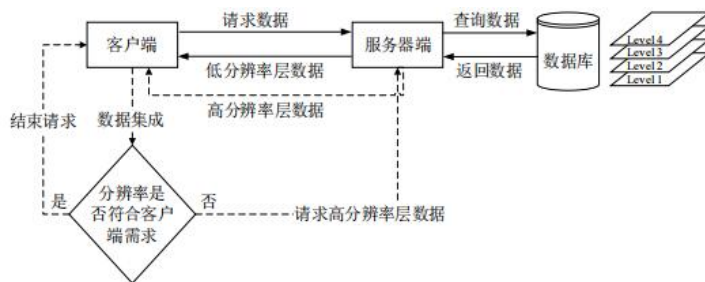


图4-10 移动GIS渐进式传输结构图

◇ **空间数据的渐进式传输可以减少用户首次访问数据时的加载时间。**

栅格影像的渐进式传输技术以视频音频等多媒体在 Internet 中传输为例, 普遍使用流媒体传输技术, **利用基本层码流提供图像轮廓信息, 增强层码流不断解码, 持续提供图像细节信息。**其关键技术为: 图像可分级编码 (SNR 可分级编码、空间分辨率可分级编码、ROI 可分级编码)。**网络 GIS 常以 JPEG 作为栅格数据传输格式。**

矢量数据的渐进式传输技术的关键在于**服务器端矢量空间数据的多尺度表达**, 主要包括两个部分: ①建立高效的 LOD 模型; ②LOD 的有效组织和管理

- ① **建立高效的 LOD 模型**: LOD 的每个模型均保留了一定层次的细节, 在绘制时, 根据不同的标准选择适当的层次模型来表示空间对象。一种方法是当地图比例尺变小时, 使用概化算子减少图面地物表达的复杂程度。

- ② **LOD 的有效组织和管理**: 采用多分辨率曲线模型保存不同层次细节。

☆ **LOD 技术**：是一种有效的图形生成加速技术，在不影响视觉效果的前提下，使用合理的算法在服务器端生成不同分辨率的矢量数据层，分别表达不同几何复杂度的地物表面细节，从而提高绘制算法效率。

移动空间数据的渐进式传输关键技术可以归纳为以下几方面：

- 1) 使用合适的算法或者编码技术生成包含不同分辨率的层次细节模型；
 - 2) 使用高效的数据结构对层次细节模型进行组织和管理；
 - 3) 移动空间数据包含丰富的空间位置信息，渐进式传输要保证拓扑关系的一致性；
 - 4) 移动 GIS 客户端接收到数据后，要能够和已有数据集成，还原细节信息。
7. **移动空间数据缓存**：在客户端和服务端建立缓存，以空间资源消耗为代价来减少网络流量和用户等待时间，提高空间数据存取效率。客户端缓存一般包括内存缓存和文件缓存；服务器端缓存包括代理缓存和服务器缓存两级。
8. **移动空间数据调用**：分为离线、在线和混合三种模式
- 1) **离线模式**：移动 GIS 的各项功能和服务不受网络条件限制，移动空间数据存储在移动终端设备上或者首次加载时从服务器端下载了完整的空间数据，移动 GIS 终端直接从本地完成数据的检索加载并进行可视化和分析。
 - 2) **在线模式**：移动 GIS 终端通过网络连接到 GIS 服务器端，实时获取移动空间数据。
 - 3) **混合模式**：将离线模式与在线模式结合起来，当网络条件良好时，移动 GIS 客户端积极向服务器端请求数据更新，同时将移动空间数据缓存到本地。当网络条件欠佳时，调用缓存于本地的数据。

➤ 移动 GIS 设计与开发以及 Android 的介绍

1. **基本原则**：①移动终端设备有效显示范围达到要求；②终端的空间数据存储容量适中；③客户端负载量不能太大；④符号显示应简洁、明了。
2. **系统结构**：
 - 1) 终端：智能手机、掌上电脑、平板电脑。
 - 2) 移动 GIS 的服务器：地理应用服务器、空间数据库
3. **WAP: Wireless Application Protocol (无线互连网应用协议)**：

应用程序和网络内容用标准格式表示，传输时用压缩编码(如紧缩二进制)，终端上使用微浏览器，以标准模式网上浏览。

WAP 使用代理技术(以 WAP 网关为其承载体)连接无线部分和 WWW。WAP 网关可为用户提供服务, 并作为网络运营管理的业务控制点。其主要功能包括:

- 1) **协议转换**: 实现 WAP 结构与 WWW 协议间转换。这些 WAP 包括: WSP、WTP、WTLS、WDP 等。
- 2) **内容编解码**: 对 WAP 数据编码, 减少数据容量和网络流量, 在终端设备上再解码, 最大程度利用无线网络传输带宽。

✧ 以中国移动为例, 使用 CMNET(China Mobile Net)可直接接入 Internet, 使用 CMWAP(China Mobile WAP)则需通过**网关**完成 WAP-WEB 协议转换后接入 Internet。

4. **Android 的体系结构**: Android 系统架构主要包括应用程序层、应用程序框架层、运行库层和 Linux 内核层

- 1) **应用程序层**: 自带的应用程序, 如短信、日历、浏览器、联系人, 以及使用者安装的第三方应用。基于 Java 开发, 开发者可通过 API 和 Android NDK(Native Development Kit)调用复杂的系统功能。
- 2) **应用程序框架层**: 对 Android 核心库和接口进一步封装, 包括活动管理器、窗口管理器、电话管理等十部分。开发者可自由访问和调用相关的 API 并发布自己的功能模块。
- 3) **运行库层**: 由一系列二进制动态库构成, 一般基于 C/C++ 实现, 通过类库提供的接口调用。Android 运行环境由核心库和 Dalvik 虚拟机组成, 前者包含框架层和应用层依赖的 Java 库, 后者采用基于寄存器的虚拟机架构设计, 负责管理对象生命周期、分配运行空间、管理堆栈、回收处理。
- 4) **Linux 内核**: 硬件和软件的中间层, Android 基于 Linux 2.6 内核开发, 主要实现对底层驱动程序的管理, 包括管理电源驱动、键盘驱动、音频驱动和 Wi-Fi 驱动等。

5. **Android 主要特点**: ①开源; ②跨平台; ③无缝集成 Google 应用。

6. **Android 程序四大基本组件**: Activity、Service、ContentProvider、BroadcastReceiver

- 1) **Activity**: 负责与用户交互, 提供可视化用户界面, 可把一个 Activity 理解成一个屏幕窗口, **生命周期有 4 种状态**:
 - a) **运行**: Activity 位于前台可见, 可以获取焦点。

- b) 暂停: Activity 可见, Activity 弹出一个对话框, 但不能获得焦点。
 - c) 停止: Activity 不可见, 失去焦点。
 - d) 销毁: Activity 结束或 Activity 所在进程结束
- 2) **Service**: Service 与 Activity 都是可执行的程序, 不同的是, Service 只能在后台运行, 没有用户界面。
- 3) **ContentProvider**: 意为广播消息接收器, 类似于编程中的事件监听器, 可以监听 Android 应用中的其他组件。
- 4) **BroadcastReceiver**: ContentProvider 为 Android 存储和读取数据提供了统一接口, 可以实现不同应用程序之间的数据共享。
7. **Android 传感器主要分为三大类**, 分别为:
- 1) **描述运动状态**: 加速度传感器、线性加速度传感器、陀螺仪传感器、重力传感器
 - 2) **描述环境状况**: 湿度传感器、光线传感器、温度传感器等;
 - 3) **描述空间位置**: 方向传感器、磁力传感器等。

➤ 移动 GIS 的应用

1. **移动电子地图**: 在移动 GIS 应用中, 往往把配置在移动终端设备上的电子地图称为移动电子地图, 是移动空间数据的一种形式。

移动电子地图的分类: 按使用移动电子地图的终端设备特点进行归类, 可将移动电子地图分为**掌上电脑、平板电脑移动电子地图和手机移动电子地图**。

移动电子地图的关键技术: 嵌入式技术、蜂窝移动网络技术、定位技术、空间数据库技术以及并发处理技术等。

2. **空间位置信息服务 (LBS)**: 【第一章概述也有提及概念】

LBS 主要由四部分组成: 客户端、网络基础设施、服务器端、地图空间数据库。

LBS 的应用软件可大致分为: 地图导航、生活服务、运动、社交、游戏等几大类。

LBS 软件平台的核心是 GIS 空间数据引擎, 其主要作用是: ①**移动定位**: 确定用户当前位置。②**信息检索**: 为用户提供与位置相关的空间信息服务和天气、出行、娱乐等生活信息服务, 以及商务信息服务、社交信息服务等。③**路径规划与引导**: 规划用户出行路径并引导其正确到达目的地。

3. **志愿者地理信息 (VGI)**: 是指在互联网时代由大量非专业用户自发创建、组织和分享的地理信息, 也叫众源地理信息。VGI 应用系统主要包括**普通大众用户、专业人员、**

通信网络、VGI 数据库和软件平台几个部分。

Goodchild 归纳了保证 VGI 数据质量的三种机制：①**众源方法**：使用集体智慧去检验和修正数据错误。②**社会方法**：设置具有更高权限的 VGI 数据管理员审核和修正数据错误。③**地理方法**：依据现有地理学理论和常识排除错误数据。

第五章 · GIS 云

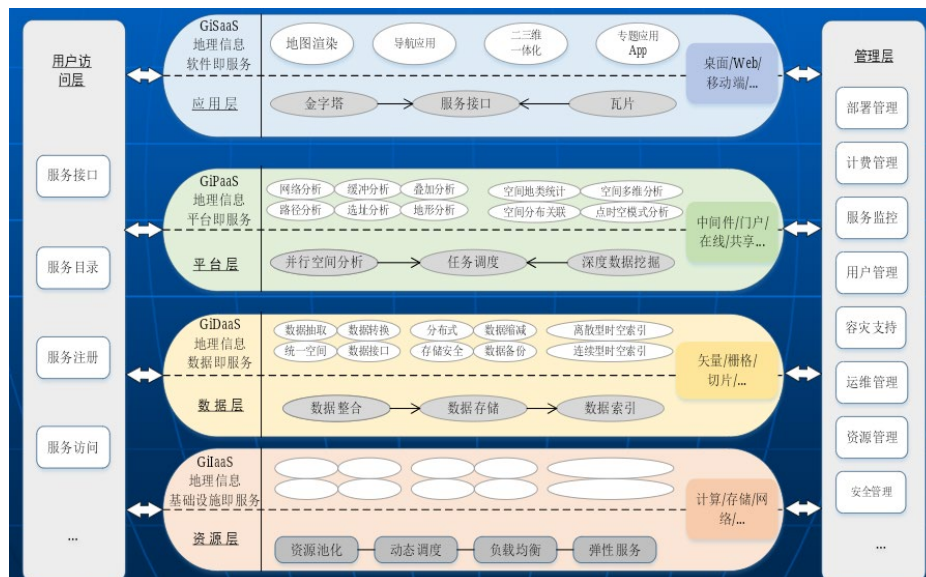
➤ 云计算

1. **云计算 (Cloud Computing)**：建立在分布式计算基础上的、实现对虚拟化的计算和存储资源池的动态部署、动态分配、实时监控的一种计算模式，以对用户透明的方式为其提供满足服务质量要求的基础架构服务、平台服务和软件服务。【云计算与网络 GIS 的结合→GIS 云 (GIS Cloud)】
2. **云计算的特点**：大规模、虚拟化、高可靠性、高通用性、高可扩展性。
3. **云安全的含义**：采取可靠的策略、技术、规范和方法来保证云计算环境中的基础设施、架构、数据、应用程序、服务等的安全。可分为三个方面：安全与隐私、规范遵循、法律与契约。**根据云计算的信息安全保障内容可将云安全分成**：
 - 1) **数据安全**：数据完整性保护、数据集中式存储隔离、数据隐私加密保护
 - 2) **网络安全**：网络层次划分、网络安全审计
 - 3) **虚拟机安全**：隔离物理机、抗旁道攻击、防火墙技术
4. **云计算的组成架构 (三横两纵)**：
 - 1) **两层管理架构**：管理层、用户访问层
 - 2) **三层服务架构 (云计算的服务模式)**：应用层、平台层、资源层
 - a) **SaaS (软件即服务)**：用户通过浏览器和 Internet 获取云上的各种软件服务的过程，是云计算的主要商业模式。
 - b) **PaaS (平台即服务)**：平台服务提供商为用户提供应用平台或开发平台等。
 - c) **IaaS (基础设施即服务)**：为用户提供计算、存储、带宽乃至服务器等基础设施服务。

➤ GIS 云基础

1. **时空大数据**：呈现出空间特征、时序特征、多维特征，并且具有时效性、海量性、复杂性、不确定性等多重特点的数据。（价值主要在时间、空间、对象之间的关联关系）
2. **GIS 云的定义**：以地理空间信息科学和分布式计算理论为基础，以地理时空大数据为处理对象，通过综合运用云计算技术和 GIS 技术解决数据密集、计算密集和 I/O 密集的应用难题，遵循时空关系和地理分布特性对地理空间数据进行存储、传输、组织、管理与分析的一种大规模高性能云服务系统。
3. GIS 云将空间数据的使用提升到复杂的**空间关联分析、时空状态拟合、智能分析决策**等更高层级。
4. **GIS 云的特点**：
 - 1) **计算资源集中管理**；
 - 2) **按需、弹性与在线服务**：
 - a) **按需服务**：按照用户要求实现服务的申请、配置和调用，服务商同时对资源及时分配和回收，以满足用户的多层次需求。
 - b) **弹性服务**：GIS 云根据服务访问规模和压力进行自动扩容、自动均衡负载。
 - c) **在线服务**：云中的各种 GIS 服务都可以在线申请、在线交付、在线使用、在线管理，地理信息服务商将各种功能和资源以在线的方式提供给用户。
 - 3) **时空大数据无缝融合**。
5. **GIS 服务器集群**：是指将多个 GIS 服务器成群组，当有用户请求到达集群服务器时，集群服务器的父节点统一对 GIS 服务进行调配，使 GIS 服务器群组发挥更高的性能、更稳定的服务以及更加灵活的服务器结构。
6. **GIS 云的部署模式**：
 - 1) **公有 GIS 云**：服务中心和数据均部署在第三方公共平台。
 - 2) **私有 GIS 云**：企业为了满足内部使用搭建的云平台，对外不提供基础设施服务。
 - 3) **混合 GIS 云**：融合了公有 GIS 云和私有 GIS 云，由多个不同的云设施以某种方式整合在一起，各个云之间相对独立。
7. **GIS 云的服务原理**：
 - 1) **四层服务架构**：资源层、数据层、平台层、应用层
 - 2) **四层服务模式**：

- a) **GIaaS 地理信息基础设施即服务** (geographic information infrastructure as a service): 企业将自己的地理信息服务部署在其他专门提供云服务的基础设施中。
 - b) **GIDaaS 地理信息数据即服务** (geographic information data as a service): 将所有地理相关信息作为服务内容向客户提供。
 - c) **GIPaaS 地理信息平台即服务** (geographic information platform as a service): 为开发人员提供 GIS 应用程序的开发环境和编程框架, 并支持以服务的形式发布。
 - d) **GISaaS 地理信息软件即服务** (geographic information software as a service): 服务提供商将 GIS 处理软件部署在自身服务器上, 客户通过浏览器和 Internet 获取这些地理信息服务。
- 3) **两种服务系统 (平台):**
- a) **GIS 数据云**: 数据管理服务及延伸、空间数据服务云。
 - b) **GIS 分析云**: 对 GIS 各种资源的虚拟化和智能部署, 提高时空大数据计算和资源调度的时效性。



8. GIS 云的关键技术: 虚拟化技术、智能化部署-服务器集群技术、资源监控技术、弹性负载均衡技术。

1) 虚拟化技术:

- a) **计算虚拟化**: CPU 虚拟化、内存虚拟化、I/O 虚拟化
- b) **存储虚拟化**: 略



- c) **网络虚拟化**：节省物理主机的网卡资源，让一个物理网络能够支持多个逻辑网络，从而高效利用网络资源。主要解决的是虚拟机构成的网络通信问题，完成的是各种网络设备的虚拟化。
- 2) **智能化部署-服务器集群技术**：实现“云”上多用户数据共享和任务协同。
- 3) **资源监控技术**：在云端实时监控系统中各类资源的使用情况和节点的承载负荷信息，及时掌握资源动态变化和节点负荷压，从而为系统资源的科学分配和调控提供必要信息。**常用的两种资源监控手段**：①**端口扫描技术**、②**Ganglia 技术**。
 - a) **端口扫描技术**：可分为 TCP 链接扫描、TCP 同步扫描、FIN 系列扫描、UDP 扫描、ACK 扫描和窗口扫描。
 - b) **Ganglia**：由 gmond、gmetad 和 gweb 三个守护进程构成。
- 4) **弹性负载均衡技术**：实时分析计算节点的负载状况，将任务调度和资源分配到各个节点，使系统稳定运行。通常包含三个步骤：①以服务调度形式处理大规模并发访问；②具有较大计算量的单个运算任务进行并行化处理；③对各个节点的负载情况进行动态分析。

➤ GIS 数据云

- 1. **GIS 数据云**：是一种特殊的云计算服务，专门针对时空大数据存储、管理与传输。
- 2. **GIS 数据云的特征**：①适应空间大数据的分布式存储与管理；②**采用多机冗余容错机制提高可靠性**；③副本机制及多路并行传输；④面向应用需求的弹性存储；⑤基于并行计算、内存计算的数据快速可视化。
- 3. **GIS 数据云的服务组织**：资源层、数据层、服务层、用户层。
 - ◇ **数据层可分为**：云获取、云存储和云索引三部分。
- 4. **GIS 云的数据服务**：时空数据整合、数据云存储服务、时空数据索引、数据备份服务、数据缩减服务、数据安全。
 - 1) **时空数据整合**：**数据预处理、异构数据整合（空间填充曲线）、数据一体化**。
 - 2) **数据云存储服务**：根据存储对象类型将数据云的分布式存储系统分为：①**分布式文件系统**、②**分布式表格系统**、③**分布式数据库系统**三种。
 - 3) **时空数据索引**：Rt-tree、3D R-tree、HR-tree、TB-tree、MVR-tree、Quad-tree 等。
 - 4) **数据备份服务**：两种有效的数据备份策略：①**副本数据部署**、②**连续数据保护**。
 - 5) **数据缩减服务**：

- a) **自动精简配置**: 利用虚拟化技术精简存储空间配置, “欺骗”操作系统, 造成存储空间足够大的假象。
 - b) **自动存储分层**: 使空间数据保留在合适的存储粗层级, 活跃数据保留在固态硬盘阵列上, 不活跃数据迁移到低速层(如磁带)上。
 - c) **重复数据删除**: 查找数据集中的重复数据, 仅保留其中一份, 消除冗余数据, 又称为消重。
 - d) **空间数据压缩**。
- 6) **数据安全**: 认证服务、加密存储、安全管理、安全日志。

➤ GIS 分析云

1. **GIS 分析云**: 利用云端强大的数据计算能力快速实现统计分析、网络分析、叠加分析等服务。并通过 REST、SOAP 等协议为前段提供按需的、安全的、可配置的地理信息服务。
2. **GIS 分析云的能力特征**: ①超强的空间分析能力; ②时空数据挖掘能力; ③便捷信息服务能力。
3. **GIS 分析云的组织服务**: 资源层、分析层、服务层、用户层。
4. **GIS 分析云的核心分析服务**: 并行空间分析、任务调度、时空数据挖掘
 - 1) **并行空间分析的核心技术体系包括**: 并行化部署和并行策略设计两部分。并行化部署的任务是实现资源的合理分配; 并行策略设计是采用并行算法模型实现资源的分布式调用和处理。

✧ **MapReduce** 是重要的并行计算模型, 将负责的计算任务进行分治计算, 并对求解后的子问题合并给出完整接。MapReduce 采用的是移动计算而非移动数据的思想, 遵循就近计算原则, 将计算移动到数据附近。MapReduce 的运行分为 Map (映射) 和 Reduce (规约) 两个阶段。设计思想见教材 P239
 - 2) **任务调度**:
 - a) **调度目标**: 计算速度快、服务质量高、系统负载均衡。
 - b) **典型调度算法**: 分为静态调度算法和动态调度算法
 - ① **Min-Min (静态)**: 优先调度预期需时最小的任务, 先易后难的贪心算法。容易造成大任务得不到响应、效率高的节点任务堆积、效率低的节点闲置的情况。

- ② **Max-Min (静态)**: 优先调度预期完成时间较长的任务, 将复杂的耗时长的任务分配给执行效率高的节点。容易造成优先级高的用户长时间得不到服务。
- ③ **Sufferage (动态)**: 通过计算每个任务在不同节点上的预期完成时间, 用次小的完成时间减去最小的完成时间, 得到不被调度情况下的值, 这个值叫做任务执行损失, 又称 Sufferage 值。优先调度 Sufferage 值大的任务, 尽可能降低总的 Sufferage 值。

任务 \ 计算节点	A	B	Sufferage值
a	2	1	1
b	4	2	2
c	6	3	3

举例: 假设a、b、c的任务量分别为2、4、6, 节点A、B的计算能力分别为1、2
 第1轮调度时, 任务c被分配给节点B执行, 任务b被分配给节点A执行;
 第2轮调度时, 任务a被分配给节点B执行(此时, c已经完成, b尚未完成)

- 3) **时空数据挖掘**: 通过地理模式识别、聚类分析、关联分析等, 在时空维度上挖掘用户感兴趣的时空模式与特征、空间非空间数据联系特征以及一些其他隐含在时空数据库中的重要特征。(三个阶段: 数据准备、数据挖掘、结果表达与解释)
- ✧ **数据仓库**: 由数据源经抽取、转换和加载过程来构造, 在多个维度合并数据, 形成部分物化的多维度数据立方体。
 - ✧ **SOLAP 立方体**: 空间数据仓库的多维模型 (定义为 $\langle Cubeid, D, M, \Gamma \rangle$)
 - ✧ **立方格**: 立方体的度量值沿着某个维度组合向上层聚集, 可以得到高层度量值, 产生的各种 GroupBy 子集。(定义为 $\langle Time, Location, LandUse \rangle$)

➤ GIS 云应用

- GEE (Google Earth Engine 谷歌地球引擎)**: 存储了近 40 年超过 200 个影像和专题产品数据集。
- ArcGIS 云**: 支持主流虚拟化环境, 同时支持用户部署在线空间数据和服务器, 实施弹性资源 GIS 服务。
 ArcGIS Online: GIS 分析云, 属于 SaaS 软件; ArcGIS Enterprise: 以 Web 为中心
- 水资源检测平台**: 以水资源的天然水循环要素的遥感监测为主要手段。